

Markergestützte Analyse gartenbaulich wichtiger Merkmale von *Phalaenopsis* Hybriden durch QTL-Kartierung und genomweite Assoziationsstudien

Von der Naturwissenschaftlichen Fakultät
der Gottfried Wilhelm Leibniz Universität Hannover
zur Erlangung des Grades einer
Doktorin der Naturwissenschaften Dr. rer. nat.

genehmigte Dissertation

von

Dipl.-Ing. agr. Inka Gawenda

geboren am 10. Mai 1978 in Neuruppin

Hannover, 2009

Referent: Herr Prof. Dr. Thomas Debener

Korreferentin: Frau Prof. Dr. Traud Winkelmann

Tag der Promotion: 28. September 2009

Inhaltsverzeichnis

1	Einleitung	1
1.1	Die Gattung <i>Phalaenopsis</i>	1
1.2	Kenntnisse über das <i>Phalaenopsis</i> Genom	6
1.3	Molekulare Marker in der Pflanzenzüchtung	6
1.4	Entwicklung von Mikrosatellitenmarkern	8
1.5	Erstellung einer genetischen Karte	9
1.6	QTL-Analysen	10
1.7	Assoziationsstudien	11
1.8	Diversitätsanalyse	12
1.9	Populationsstruktur	13
1.10	Kopplungsungleichgewicht	14
1.11	Ziel der Arbeit	16
2	Material und Methoden	17
2.1	Pflanzenmaterial	17
2.2	Sonstige Materialien	18
2.3	Aufnahme der phänotypischen Daten	18
2.4	Molekularbiologische Methoden	19
2.4.1	DNA-Extraktion	19
2.4.2	Polyacrylamid Gelelektrophorese	20
2.4.3	Herstellung einer für Mikrosatelliten angereicherten DNA Bank	21
2.4.4	Testen und Anwendung der Mikrosatelliten-Primerpaare	25
2.4.5	Anpassung und Anwendung der AFLP-Technik	27
2.5	Bioinformatik	29
2.5.1	Screening der Sequenzen nach Mikrosatellitenmotiven	29
2.5.2	Design von Primern	29
2.5.3	Erstellung der genetischen Karte	29
2.5.4	QTL-Kartierung	31
2.5.5	Phylogenetische Analysen	32

2.5.6	Bestimmung/Ermittlung der Populationsstruktur	32
2.5.6.1	Berechnung der Anzahl der Subpopulationen nach Pritchard et al. (2000)	33
2.5.6.2	Berechnung der Anzahl der Subpopulationen nach Cockram et al. (2008)	34
2.5.6.3	Berechnung der Anzahl der Subpopulationen nach Evanno et al. (2005)	34
2.5.7	Ermittlung des Kopplungsungleichgewichts	35
2.5.8	Berechnung der Marker-Eigenschafts-Assoziationen	35
2.5.8.1	Assoziationsstudie in R	35
2.5.8.2	Assoziationsstudien in TASSEL 2.1	36
2.5.9	Graphische Darstellung	36
3	Ergebnisse	37
3.1	Kartierung gartenbaulich wichtiger Merkmale	37
3.1.1	Phänotypische Merkmale	37
3.1.1.1	Zusammenhang zwischen phänotypischen Merkmalen	41
3.1.2	Mikrosatellitenmarker	42
3.1.2.1	Entwicklung von Mikrosatellitenmarkern	42
3.1.2.2	Anwendung der Mikrosatellitenmarker	44
3.1.3	Erstellung der genetischen Karte	44
3.1.3.1	Paternale Karte	45
3.1.3.2	Maternale Karte	47
3.1.4	QTL-Analyse in der Nachkommenschaft	49
3.2	Genomweite Assoziationsstudie	57
3.2.1	Diversitätsanalyse	57
3.2.1.1	Variationskoeffizient	57
3.2.1.2	Phylogenetischer Baum	58
3.2.2	Populationsstruktur	60
3.2.2.1	Anzahl an Subpopulationen	60
3.2.2.2	Ermittlung der Anzahl an Subpopulationen nach Pritchard et al. (2000)	62
3.2.2.3	Ermittlung der Anzahl an Subpopulationen nach Cockram et al. (2008)	64
3.2.2.4	Ermittlung der Anzahl an Subpopulationen nach Evanno et al. (2005)	67
3.2.3	Kopplungsungleichgewicht zwischen den Markern	69

3.2.4	Berechnungen der Marker-Eigenschafts-Assoziationen	73
3.2.4.1	Assoziationsstudie mit R 2.8.1	73
3.2.4.2	Assoziationsstudie mit TASSEL	75
3.2.4.3	Vergleich der Ergebnisse der Assoziationsstudie berechnet in R 2.8.1 und TASSEL 2.1	76
3.3	Verknüpfung der familienbasierten QTL-Analyse und der genomweiten Assoziationsstudie	78
4	Diskussion	81
4.1	Kartierung gartenbaulich wichtiger Merkmale	81
4.1.1	Phänotypische Merkmale in der Nachkommenschaft	81
4.1.2	Molekulare Marker	82
4.1.2.1	Entwicklung von Mikrosatellitenmarkern	82
4.1.2.2	Anpassung des AFLP-Markersystems	85
4.1.3	Erstellung der genetischen Karte	85
4.1.4	QTL-Analyse	88
4.2	Assoziationsstudie	91
4.2.1	Phänotypische Daten in der Genotypensammlung	92
4.2.2	Diversitätsanalyse	92
4.2.3	Populationsstruktur	96
4.2.4	Vergleich Diversitätsanalyse und Ermittlung der Populationsstruktur	98
4.2.5	Kopplungsungleichgewicht	99
4.2.6	Berechnung der Marker-Eigenschafts-Assoziationen	102
4.3	Vergleich familienbasierte QTL-Analyse und genomweite Assoziationsstudie	105
4.4	Ausblick	107
5	Anhang	125
5.1	Zusätzlich verwendetes Material	125
5.1.1	Bakterienstamm	125
5.1.2	Verwendete Geräte	126
5.1.3	Hersteller der Geräte und Chemikalien	127
5.1.4	Zusammensetzung von Medien, Puffer und Lösungen	128
5.2	Mikrosatellitenmarker	131
5.3	Oligonukleotide für die AFLP-Technik	134
5.4	Signifikanzschwellen für das Intervallmapping	136
5.5	Ermittlung der Anzahl der Subpopulationen nach Evanno et al. (2005)	138

5.6	R Routine	139
5.7	Visuelle Darstellung der Ergebnisse aus R 2.8.1	142
5.8	Genotypensammlung	146

Abkürzungsverzeichnis

<i>AFLP</i>	Amplified Fragment Length Polymorphism
<i>bp</i>	Basenpaare
<i>cM</i>	centi Morgan
<i>EST</i>	Expressed Sequence Tag
<i>FIASCO</i>	Fast Isolation by AFLP of Sequences Containing repeats
<i>GC</i>	Genomic Control
<i>GLM</i>	General Linear Model
<i>IPTG</i>	Isopropyl- β -D-thiogalactopyranosid
<i>IRD</i>	Infra Red Dye
<i>LD</i>	Linkage Disequilibrium = Kopplungsungleichgewicht
<i>LG = KG</i>	Linkage Group = Kopplungsgruppe
<i>LOD</i>	Logarithm of the Odds
<i>MCMC</i>	Monte Carlo Markov Chain
<i>PAA</i>	Polyacrylamid
<i>PIC</i>	Polymorphism Information Content
<i>QTL</i>	Quantitative Trait Loci
<i>RFLP</i>	Restriktion Fragment Length Polymorphism
<i>SAMPL</i>	Selective Amplification of Microsatellite Polymorphic Loci
<i>SCL</i>	Strongest Cross Link
<i>SDM</i>	Single Dose Marker
<i>SDRF</i>	Single Dose Restriktion Fragment
<i>SNP</i>	Single Nucleotide Polymorphism
<i>SSR</i>	Simple Sequence Repeat
<i>X - Gal</i>	5-Brom-4-Chlor-3-Indoxyl- β -D-Galactopyranosid

Abbildungsverzeichnis

1.1	Verbreitungsgebiet der Gattung <i>Phalaenopsis</i>	2
1.2	Abfall des Kopplungsungleichgewichts in Abhängigkeit von der Zeit/Anzahl der Generationen	16
2.1	Pflanzenmaterial	18
2.2	Amplifikationsschema der M13-Mikrosatelliten PCR	26
3.1	Variation der Blütenmorphologie in der Nachkommenschaft '21x19' .	38
3.2	Unterschiede in der Ausprägung des Anthocyangehalts der Blüten in der Nachkommenschaft '21x19'	38
3.3	Unterschiede in Blütengröße und Rispenlänge	39
3.4	Histogramme	40
3.5	Zusammenhang zwischen phänotypischen Eigenschaften	41
3.6	Übersicht der Kopplungsgruppen der paternalen Karte	46
3.7	Übersicht der Kopplungsgruppen der maternalen Karte	48
3.8	QTL für Blütengröße	51
3.9	QTL für Anthocyangehalt in der Lippe	52
3.10	QTL für Anthocyangehalt in der Blüte	53
3.11	QTL für Anthocyangehalt im Blatt	54
3.12	QTL für Rispenlänge (Teil 1)	55
3.13	QTL für Rispenlänge (Teil 2)	56
3.14	Histogramm der Variationskoeffizienten	58
3.15	Neighbour-Joining Dendrogramm	59
3.16	Populationsstruktur mit K=10 Subpopulationen	61
3.17	Populationsstruktur mit K=11 Subpopulationen	61
3.18	Populationsstruktur mit K=12 Subpopulationen	62
3.19	Gegenüberstellung des Dendrogramms mit den einzelnen Subpopula- tionen	63
3.20	Graphische Darstellung der lnP(D)-Werte aus STRUCTURE 2.2 in Abhängigkeit der Anzahl an Subpopulationen (K)	65
3.21	Graphische Darstellung der Korrelation zwischen den Q Matrizen . .	66

3.22	Ermittlung der Anzahl der Subpopulationen nach Evanno et al. (2005)	68
3.23	Kopplungsungleichgewicht (LD) in Abhängigkeit vom Abstand in cM	69
3.24	Kopplungsungleichgewicht zwischen kartierten Markern	70
3.25	Kopplungsungleichgewicht zwischen AFLP-Markern	71
3.26	Kopplungsungleichgewicht zwischen Mikrosatellitenmarkern	72
3.27	Assoziationen mit der Blütengröße	74
3.28	QTL und Assoziationen für die Eigenschaft Blütengröße	79
3.29	QTL und Assoziationen für die Eigenschaft Rispenlänge	80
4.1	Vergleich familienbasierte Kopplungsanalyse und Assoziationsstudie .	106
4.2	Übersicht der Markerentwicklung für die Markergestützte Selektion .	107
5.1	Neighbour-joining Dendrogramm basierend auf sechs Mikrosatellitenmarkern	131
5.2	Ermittlung der Anzahl der Subpopulationen nach Evanno et al. (2005)	138
5.3	Ergebnisse aus R 2.8.1 für Anthocyangehalt im Blatt	142
5.4	Ergebnisse aus R 2.8.1 für Blütenfarbe	142
5.5	Ergebnisse aus R 2.8.1 für Blütenkonsistenz	143
5.6	Ergebnisse aus R 2.8.1 für Blütengröße	143
5.7	Ergebnisse aus R 2.8.1 für Blütentyp	143
5.8	Ergebnisse aus R 2.8.1 für Duft	144
5.9	Ergebnisse aus R 2.8.1 für Blattmuster	144
5.10	Ergebnisse aus R 2.8.1 für Blattform	144
5.11	Ergebnisse aus R 2.8.1 für Blatttyp	145
5.12	Ergebnisse aus R 2.8.1 für Herkunft	145
5.13	Ergebnisse aus R 2.8.1 für Rispenlänge	145

Tabellenverzeichnis

1.1	Die bedeutensten Schritte in der Geschichte der <i>Phalaenopsis</i> Züchtung	5
2.1	Phänotypischen Merkmale in der Genotypensammlung	19
2.2	Eingesetzte Mengen für die Hybridisierung der biotinylierten Sonden an die DNA-Fragmente	21
2.3	Waschschritte zum Entfernen nicht gebundener Fragmente	22
2.4	Reaktionsbedingungen Insert PCR	24
2.5	M13-Mikrosatelliten PCR	26
2.6	Reaktionsbedingungen der AFLP Präamplifikation	28
2.7	PCR Bedingungen für AFLP-Endreaktion	28
3.1	Kolmogorov-Smirnov-Test der phänotypischen Daten	37
3.2	Entwicklung von Mikrosatellitenmarker	43
3.4	Mikrosatellitenmarker und ihre PIC-Werte	44
3.5	Übersicht der für die Kartierung verwendeten Marker	45
3.6	Ankermarker	47
3.7	QTL in der Nachkommenschaft	50
3.8	Ermittlung der Anzahl der Subpopulationen	64
3.9	Zusammenfassung: Assoziationsstudie mit R 2.8.1	75
3.10	Zusammenfassung: Assoziationsstudie mit TASSEL 2.1	76
3.11	Signifikant assoziierte Marker	77
4.1	Faktoren, die das Kopplungsungleichgewicht in einer Population be- einflussen	101
5.1	In der vorliegenden Arbeit verwendete Bakterienstamm	125
5.2	Verwendete Geräte	126
5.3	Standorte der Hersteller	127
5.4	Medien	128
5.5	Puffer und Lösungen	129
5.6	Internetadressen der frei verfügbaren Software	130
5.7	Mikrosatellitenprimerpaare entwickelt aus dem Genotyp 0744	132

5.9	veröffentlichte Mikrosatellitenprimerpaare	133
5.11	AFLP-Adaptoren	134
5.12	AFLP-Oligonukleotide	134
5.13	AFLP-Primerkombinationen	135
5.14	Signifikanzschwellen für das Intervallmapping	136
5.15	Gruppe 4	146
5.16	Gruppe 1A	147
5.17	Gruppe 2A	148
5.18	Gruppe 3	149
5.19	Gruppe 2C	150
5.20	Gruppe 2B (Teil 1)	151
5.21	Gruppe 2B (Teil 2)	152
5.22	Gruppe 1C (Teil 1)	153
5.23	Gruppe 1B (Teil 2)	154

Zusammenfassung

Phalaenopsis Hybriden gehören zu den wichtigsten Zierpflanzen weltweit. Trotz der hohen ökonomischen Bedeutung ist jedoch kaum etwas über die genomische Struktur bekannt. In dieser Studie wurden 17 neue Mikrosatellitenmarker entwickelt und das AFLP-Markersystem an das *Phalaenopsis*-Genom angepasst. Mit Hilfe dieser Marker wurde die erste genetische Karte für *Phalaenopsis* erstellt. Obwohl diese Karte nur ca. 30% des Genoms abdeckt, konnten bereits QTL für die Merkmale 'Blütengröße' und 'Anthocyangehalt in der Blüte' detektiert werden.

Parallel zur familienbasierten Kopplungsanalyse sollte eine genomweite Assoziationsstudie mit einer Population, bestehend aus 134 *Phalaenopsis* Hybriden und 10 Wildarten durchgeführt werden. Die verwendete Population stellt eine repräsentative Stichprobe des heute auf dem europäischen Markt befindlichen Sortiments dar. Mit Hilfe der Diversitätsanalyse und der Ermittlung der Populationsstruktur kann die Züchtungsgeschichte von *Phalaenopsis* Hybriden in Europa nachvollzogen werden. Die Distanzen wurden in PHYLIP 3.68 nach Nei and Li (1979) berechnet und ein Neighbour-Joining Baum nach Saitou and Nei (1987) erstellt. Die Populationsstruktur wurde mit STRUCTURE 2.2 ermittelt. Um die Anzahl an Subpopulationen zu ermitteln, wurden drei Methoden (nach Pritchard et al. (2000), Cockram et al. (2008) und Evanno et al. (2005)) angewandt. Es konnten 10 bis 12 Subpopulationen ermittelt werden.

Eine weitere Voraussetzung zur Durchführung einer genomweiten Assoziationsstudie ist die Kenntnis über das Kopplungsungleichgewicht in einer Population. Die hier durchgeführten Analysen zeigen einen relativ langsamen Abfall des Kopplungsungleichgewichts in Abhängigkeit der genetischen Distanzen zwischen den Markern. Aufgrund dessen konnten durch eine erste genomweite Assoziationsstudie mit nur relativ wenig Markern Assoziationen zwischen Markern und gartenbaulich wichtigen Merkmalen wie z.B. Rispenlänge und Blütengröße ermittelt werden.

Die hier durchgeführte Studie bietet einen ersten Überblick über die genomische Struktur von *Phalaenopsis* Hybriden. Es wurde gezeigt, dass Assoziationsstudien eine geeignete Methode zur Detektion von gartenbaulich wichtigen Eigenschaften in *Phalaenopsis* sind. Auf Grundlage dieser Arbeit können weitere Analysen zur Validierung der hier gefundenen QTL durchgeführt werden, welche anschließend zur verbesserten Züchtung neuer Sorten mittels markergestützter Selektion verwendet werden können.

Schlagwörter: Kopplungskarte – Populationsstruktur – Kopplungsungleichgewicht

Abstract

Phalaenopsis hybrids belong to the most important ornamental pot plants worldwide. Despite its high economical value, there is hardly anything known about the genomic structure. At this study 17 new microsatellite markers were developed and the AFLP-marker technique was adapted to the *Phalaenopsis* genome. The first genetic map for *Phalaenopsis* was constructed with these markers. Notwithstanding this map covers only approx. 30% of the genome, QTLs for 'flower size' and 'anthocyanine content in the flower' could be detected.

In parallel to family based linkage analysis a genomewide association study was conducted with a population comprising 134 *Phalaenopsis* hybrids and 10 wild species. This population represents an adequate sample of the European *Phalaenopsis* assortment. Due to analysis of genetic diversity and revealing the population structure, the European breeding history of *Phalaenopsis* Hybrids could be reconstructed. Genetic distances were calculated with PHYLIP 3.68 according to Nei and Li (1979) and a neighbour-joining tree was constructed according to Saitou and Nei (1987). Population structure was revealed by STRUCTURE 2.2 and the number of subpopulations were identified by three different strategies (Pritchard et al. (2000), Cockram et al. (2008) and Evanno et al. (2005)). Ten to twelve subpopulations could be determined.

Another prerequisite for genomewide association analysis is the determination of linkage disequilibrium between markers in a population. In this study a slow decay of linkage disequilibrium of markers subjected to their genetic distance could be obtained. Accordingly, marker-trait associations (f.e. flower size and panicle length) could be detected with comparatively few markers.

This study gives an overview of the genomic structure of *Phalaenopsis* hybrids. It is shown, that genomewide association studies are an appropriate method to detect horticulturally important traits for *Phalaenopsis*. Based on these results further analysis can be done to validate the QTLs and marker-trait-associations, that can be used to improve *Phalaenopsis* breeding by marker assisted selection.

Keywords: Linkage map – Population structure – Linkage disequilibrium

1 Einleitung

1.1 Die Gattung *Phalaenopsis*

Die Gattung *Phalaenopsis* gehört zu einer der größten Pflanzenfamilien, der Familie *Orchidaceae*, und umfasst laut Christenson (2001) ungefähr 66 Arten. Im Gegensatz zu früheren Einteilungen, zum Beispiel von Sweet (1980), bezieht Christenson auch *Doritis* ssp. und *Kingidium* ssp. in die Gattung *Phalaenopsis* ein. Des Weiteren unterteilt Christenson die Gattung *Phalaenopsis* in fünf Untergattungen: *Proboscidioides*, *Aphyllae*, *Parishianae*, *Polychilos* und *Phalaenopsis*. Das ursprüngliche Herkunftsgebiet erstreckt sich von Ostindien, Südchina und Taiwan über Indonesien den Philippinen und Papua-Neuguinea bis in den Norden Australiens. Das Verbreitungsgebiet ist in Abbildung 1.1 dargestellt.

Gruss and Wolff (1995) fassen die phänotypischen Merkmale und Eigenschaften von *Phalaenopsis* Arten folgendermaßen zusammen: Meist wachsen die Pflanzen epiphytisch auf Bäumen, wobei die Wurzeln neben der Nährstoffaufnahme auch zur Verankerung dienen. Da die Gattung *Phalaenopsis* zu den monopodialen Orchideen gehören, bilden sie keine Seitensprosse sondern wachsen nur nach oben. Die Blätter wachsen gegenständig wobei alle 6-12 Monate ein neues meist mehr oder weniger ledriges Blatt gebildet wird. Zwischen den Blättern entsteht der Blütenstand, an dem die Blüten in Form einer Traube oder Rispe angeordnet sind. Die Blüten bestehen aus drei Petalen und drei Sepalen, wobei die untere Sepale meist stark zu der so genannten Lippe verformt ist. Aufgrund der enormen Ähnlichkeit der Petalen und Sepalen bei den Monocotyledonen werden sie auch zusammen als Tepalen bezeichnet (Bowman (1997)). Die Blütengröße variiert von ca. 3 bis ca. 12 cm im Durchmesser. Die Blütenfarben reichen von weiß, gelb und leicht grün bis zu rosa, pink, violett und rot mit zum Teil sehr unterschiedlichen Zeichnungen. Die Lippe oder auch Labellum genannt, ist meist wesentlich kräftiger gefärbt. Die männlichen und weiblichen Organe sind in der Mitte der Blüte als so genannte Columna oder Gynostemium angeordnet. Die Antheren enthalten zwei bis vier kugelförmige Pollinien, in denen die Pollen enthalten sind. Nach der Befruchtung werden Kapseln gebildet, in denen innerhalb von fünf bis fünfzehn Monaten bis zu zehntausend Samen heranreifen.

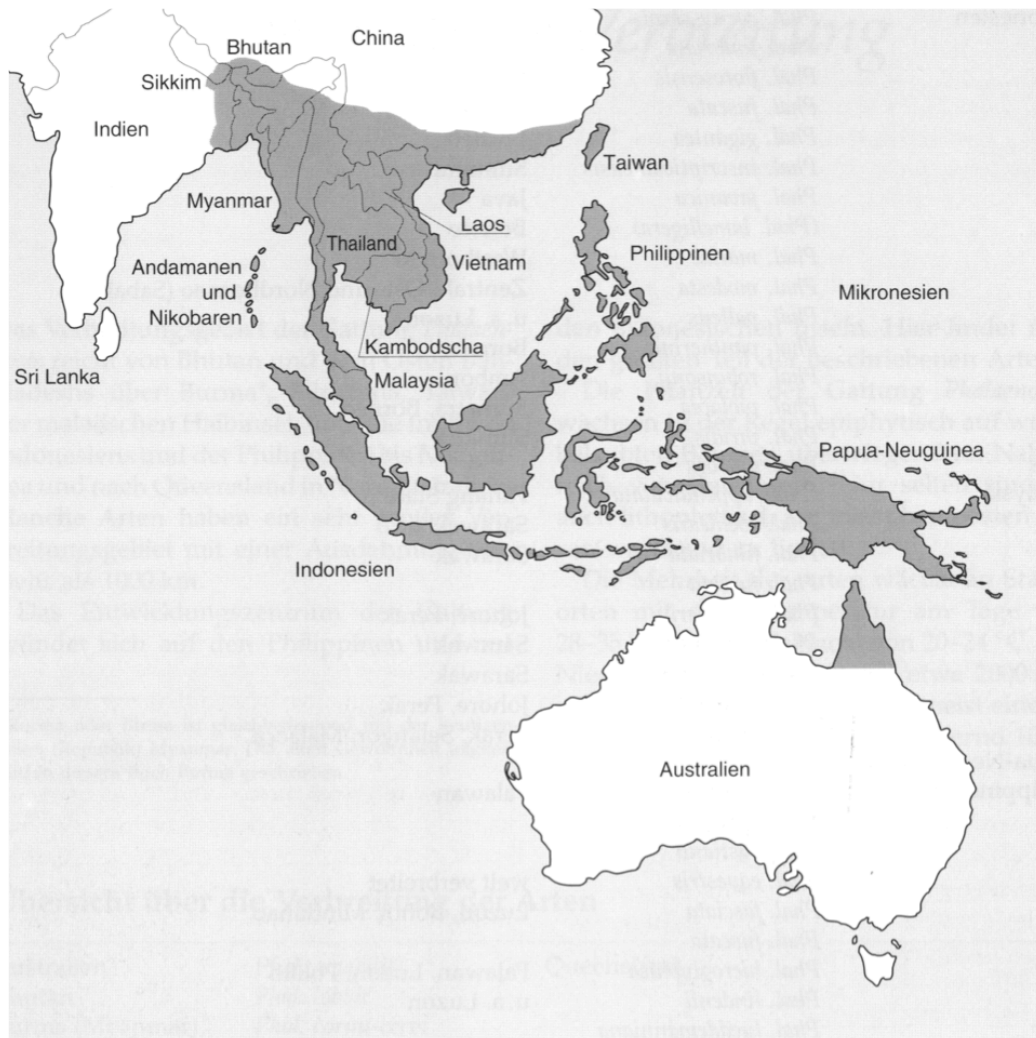


Abbildung 1.1: Verbreitungsgebiet der Gattung *Phalaenopsis* (aus: Gruss and Wolff (1995))

Gegen Ende des 19. Jahrhunderts und Anfang des letzten Jahrhunderts wurden die ersten Hybriden gezüchtet. Die ersten *Phalaenopsis* Hybriden gehen auf Kreuzungen mit *Phalaenopsis amabilis* zurück und hatten vor allem große, lang haltbare Blüten als Zuchtziel (Griesbach (1985)). Interspezifische Kreuzungen und die Verdoppelung der Chromosomen sind Techniken, die zur Züchtung neuer Hybriden angewandt werden (Arditti (1992), Griesbach (1985)). Allerdings ist die generative Vermehrung von *Phalaenopsis* sehr langsam und damit auch relativ teuer. Aus diesem Grunde wurden die ersten Züchtungen von wenigen reichen Adligen in England, Frankreich und Deutschland vorgenommen (Christenson (2001)). Des Weiteren waren die ersten *in vitro* Techniken aufgrund von zu großer somaklonaler Variation weniger geeignet für die Massenproduktion von *Phalaenopsis* Hybriden (Griesbach (2002)). Erst seit den 90er Jahren des letzten Jahrhunderts, mit der Verbesserung und damit dem Beginn der kommerziellen *in vitro* Vermehrung, war es möglich, *Phalaenopsis* kostengünstig als uniforme Klone zu vermehren. Die Produktion stieg von ca. 5 Millionen Pflanzen im Jahr 1998 auf über 31 Millionen Pflanzen im Jahr 2004 (Winkelmann et al. (2006)) und fast 49 Millionen Pflanzen im Jahr 2007 allein in Europa (Heinrichs (2008)). Mit einem Umsatzanteil von 24% lag die Beliebtheit von *Phalaenopsis* Topf-Orchideen in Deutschland im Jahr 2008 weit vor *Cyclamen* und *Poinsettia* mit jeweils 8% Umsatzanteile (Niehues (2009)). In Tabelle 1.1 sind die bedeutendsten Schritte in der Züchtung von *Phalaenopsis* Hybriden aufgeführt. Unter anderem wird deutlich, dass neben technischen Entwicklungen, wie die Verbesserung der *in vitro* Technik, auch die Präferenzen der Verbraucher einen enormen Einfluss auf die Züchtung haben.

Da es die heutige *in vitro*-Technik ermöglicht, *Phalaenopsis* Hybriden in großer Stückzahl vegetativ mit hoher Homogenität zu vermehren, ist dem Züchter Homogenität in der Kreuzung weniger wichtig, als neue Sorten mit verbesserten Eigenschaften zu kreieren. Aktuelle Zuchtziele sind im folgenden aufgeführt (F. von Rundstedt, Bock Bio Science, persönliche Mitteilung):

- Schnellwüchsigkeit
- Mehrrispigkeit (mindestens zwei Rispen pro Pflanze)
- Verzweigung der Rispe
- Stärke der Rispe
- Anzahl der Blüten pro Rispe (mindestens 7-9)
- aufrecht stehendes Laub (dies vereinfacht eine Automatisierung in der Produktion)

- Blütenfarbe
- Blütenform
- Blütengröße
- Rispenlänge (55-60 cm werden für 12 cm Topf präferiert)
- Haltbarkeit der Blüten
- kein Knospenfall
- Resistenz/Toleranz gegen z.B. Virosen, Bakteriosen, pilzliche Erkrankungen, *Rizoctonia*, *Fusarium*
- keine Vorblüher
- keine Luftwurzeln
- gemustertes Laub
- Duft

Parallel zur steigenden Beliebtheit von *Phalaenopsis* Hybriden wurden auch immer mehr neue Hybriden gezüchtet. So wurden in den Jahren von 1976 bis 2008 durchschnittlich 100 neue *Phalaenopsis* Hybriden von der Royal Horticulture Society alle zwei Monate registriert (Chang et al. (2009)). Bei genauerer Betrachtung der Züchtungsgeschichte von *Phalaenopsis* Hybriden ist jedoch zu sehen, dass die heutigen Hybriden auf nur wenige Arten zurückzuführen sind. Gruss and Wolff (1995) zeigen auf, dass viele Hybriden einen neuen Namen erhielten und somit irrtümlich als neue, eigenständige Sorte betrachtet werden. Genetisch basieren die heutigen Hybriden größtenteils auf wenigen Wildarten wie *Ph. aphrodite*, *Ph. equestris*, *Ph. amabilis*, *Ph. violaceae*, *Ph. schilleriana* und *Ph. sanderiana*.

Tabelle 1.1: Die bedeutensten Schritte in der Geschichte der *Phalaenopsis* Züchtung;
Zusammenfassung aus Griesbach (2002)

1875-1886	Die erste Hybride (<i>Ph.</i> 'Intermedia') entsteht in England als Kreuzung zwischen <i>Ph. amabilis</i> und <i>Ph. equestris</i> . Später wird diese Hybride auch als natürliche Hybride gefunden.
1909	H. Burgeff (Dtl.) und N. Bernard (Frkr.) entdecken, dass pilzliche Infektion zur Keimung der Samen nötig ist
Anfang 1920er	<i>Phalaenopsis</i> als Schnittblume gewinnt an Bedeutung
1920	Die Hybride <i>Ph.</i> 'Gilles Gratiot', eine Kreuzung aus <i>Ph. amabilis</i> und der <i>rimestandiana</i> ¹ -Form wurde registriert,
1922	L. Knudson entwickelt 'Knudson C' Medium, welches nicht symbiotische Keimung ermöglicht
1927	Die Hybride <i>Ph.</i> 'Elisabethae' mit der Blütenkonsistenz von <i>rimestandiana</i> ¹ und der Blütengröße der <i>grandiflora</i> ¹ -Form wurde registriert
1940	<i>Ph.</i> 'Doris' wurde registriert (erste tetraploide, rein weiße, künstliche Hybride mit großen festen Blüten)
1950er	pinke Hybriden werden auf Grundlage von <i>Ph. schilleriana</i> , <i>Ph. sanderiana</i> und <i>Ph. amabilis</i> bzw. <i>Ph.</i> 'Doris' gezüchtet
1960er	Kunden präferieren <i>Phalaenopsis</i> wieder mehr als Topfpflanze; 'Neue' Hybriden entstehen durch interspezifische Kreuzungen mehrerer Arten. Viele experimentelle Kreuzungen werden durchgeführt. Farbige und gemusterte Blüten werden gezüchtet. Poliploide Formen treten vermehrt auf.
1980er	<i>Phalaenopsis</i> Hybriden werden beliebteste Orchidee. Produktion und Züchtung werden auf Massenproduktion ausgerichtet, d.h. Eigenschaften wie Uniformität werden in der Züchtung genauso wichtig wie Eigenschaften, die den Zierwert betreffen.
1971-1996	Verbesserung der <i>in vitro</i> Technik und damit einhergehend Umstellung der Massenvermehrung auf vegetative <i>in vitro</i> Produktion

¹ verschiedene Formen von *Ph. amabilis* mit größerer Blüte (*grandiflora*) und festerer Konsistenz (*rimestandiana*)

1.2 Kenntnisse über das *Phalaenopsis* Genom

Trotz der enorm hohen ökonomischen Bedeutung von *Phalaenopsis* Hybriden für den Zierpflanzenbereich, wurden bisher nur wenige molekulare Untersuchungen des Genoms durchgeführt. Viele wissenschaftliche Arbeiten konzentrierten sich auf die Verbesserung der *in vitro* Produktion von *Phalaenopsis* (z.B. Gow et al. (2008), Shrestha et al. (2007)), Gentechnologie (z.B. Chai et al. (2007), Liao et al. (2004)) und die Detektion viraler Pathogene (z.B. Ajjikuttira et al. (2002), Eun et al. (2000)) sowie die molekulare Analyse von Blütenmerkmalen (z.B. Chen et al. (2007), Guo et al. (2007), Ma et al. (2009), Tsai et al. (2008), Tsai (2005)). Phylogenetische Untersuchungen an *Phalaenopsis* wurden zunächst nur durch Betrachtung eines Teils des Genoms wie zum Beispiel dem Chloroplastengenom (Kim et al. (2003)) oder der ITS-Region der ribosomalen DNA (Tsai et al. (2005)) durchgeführt. Des Weiteren wurden bei diesen Untersuchungen vorrangig *Phalaenopsis* Wildarten untersucht. Chang et al. (2009) veröffentlichten erstmals eine Untersuchung, welche *Phalaenopsis* Hybriden einschloss und das Gesamtgenom mit Hilfe von AFLP-Markern betrachtete.

Die Chromosomen von *Phalaenopsis* sp. und Hybriden haben eine Größe von 1,5-3,5 μm (Arends (1970)). Der DNA Gehalt und die Ploidiestufe von verschiedenen *Phalaenopsis* Wildarten wurde unter anderem von Bennett (2005) sowie von Lin et al. (2001) untersucht. Dabei stellten sie fest, dass die DNA in 19 Chromosomen organisiert ist und ihr Gehalt zwischen 1,18 pg ($\sim 1,15 \times 10^9$ bp) und 8,31 pg ($\sim 8,13 \times 10^9$ bp) mit einem Mittelwert von 3,67 pg ($\sim 3,59 \times 10^9$ bp) liegt. Kurz nach der Etablierung der *in vitro* Technik zur *Phalaenopsis* Vermehrung wurde die Züchtung entscheidend durch den tetraploiden Sämling *Ph.* 'Doris' verändert (Christenson (2001)). Da Polyploidie die genetische Variabilität erhöhen und die Blütengöße sowie -konsistenz positiv beeinflussen kann, wurde in *Phalaenopsis* unbewusst auf Polyploide selektiert. Durch Kreuzung farbiger Arten mit modernen autotetraploiden, weißen Arten wurden in den 1960ern farbige Hybriden mit einem Ploidielevel bis zu hexaploid gezüchtet, wodurch die Haltbarkeit, Blütengröße sowie -konsistenz erhöht werden konnte (Griesbach (1985)). Interspezifische Kreuzungen und die Verdopplung des Chromosomensatzes sind beliebte Techniken in der *Phalaenopsis*züchtung. Heute wird davon ausgegangen, dass die meisten *Phalaenopsis* Hybriden tetraploid sind (Griesbach (1985)).

1.3 Molekulare Marker in der Pflanzenzüchtung

Die Grundlagen der gezielten wissenschaftlichen Pflanzenzüchtung bildeten die Arbeiten von Charles Darwin und Gregor Mendel. Mit der Entwicklung neuer Methoden wie Biotechnologie, Genomforschung und Anwendung von molekularen Markern, wurde

gegen Ende des letzten Jahrhunderts die Molekulare Pflanzenzüchtung als interdisziplinäre Wissenschaft begründet (Moose and Mumm (2008)). In der Pflanzenzüchtung wurden Zuchtstrategien anfangs auf Grundlage phänotypischer Merkmale getroffen (Bernardo (2008)). Die phänotypischen Merkmale bzw. Marker sind im Gegensatz zu molekularen Markern jedoch stark umweltabhängig. Molekulare Marker können mit phänotypischen Eigenschaften korrelieren und bieten die Vorteile, dass sie stabil und in allen Teilen der Pflanze, egal welchen Alters, erfasst werden können. Sie werden weder von der Umwelt noch von pleiotrophen oder epistatischen Effekten beeinflusst (Agarwal et al. (2008)).

In den letzten zwei Jahrzehnten wurden eine Reihe von Techniken zur Detektion molekularer Marker entwickelt. Eine Übersicht der wichtigsten Techniken wird von Agarwal et al. (2008) gegeben. Es wird zwischen PCR-basierten und nicht PCR-basierten Techniken unterschieden. Alle Markertechniken haben Vor- und Nachteile, die eine Kombination verschiedener Techniken zur Analyse einer Gattung oder Fragestellung befürworten. In der vorliegenden Arbeit wurden AFLP- (**A**mplified **F**ragment **L**ength **P**olymorphism) und Mikrosatellitenmarker zur Analyse der genomischen Struktur von *Phalaenopsis* Hybriden verwendet. Beide Markertypen sind gut untersucht und werden routinemäßig für genetische Studien verwendet.

Die AFLP Marker Technologie wurde erstmals von Vos et al. (1995) veröffentlicht. Die Methode basiert auf einer Restriktion der DNA mit zwei verschiedenen Enzymen und der selektiven Amplifikation der DNA-Fragmente. Je nach Anzahl der selektiven Nukleotide, der Primerkombination, dem Motiv der selektiven Nukleotide, dem GC-Gehalt und der Genomgröße und -komplexität, können 50-100 AFLP-Banden je Probe erzeugt werden (Agarwal et al. (2008)). Die Technik ist relativ leicht automatisierbar und wird heute für viele genetische Analysen, wie beispielsweise die Erstellung von genetischen Karten (Barcaccia et al. (2003)), Diversitätsanalysen (Chang et al. (2009), Gawenda and Debener (2009)), evolutionäre Studien und genomweite Assoziationsstudien (Achleitner et al. (2008)) angewandt. Sie kann zur Erstellung hochauflösender Karten beitragen, welche zur Positionsklonierung und Identifizierung von Kandidatengenen genutzt werden können (Vuylsteke et al. (2007)). Damit bietet die AFLP-Technik eine relativ schnelle und kostengünstige Methode, um viele Genotypen mit einer hohen Auflösung und ohne vorherige Kenntnisse über das Genom zu untersuchen (Meudt and Clarke (2007)).

Mikrosatellitenmarker, oder auch SSR-Marker (**S**imple **S**equences **R**epeats) genannt, sind kurze sich wiederholende Sequenzen, die im Genom aller bisher untersuchten Prokaryoten und Eukaryoten vorkommen (Zane et al. (2002)). Aufgrund der hoch konservierten flankierenden Regionen werden Mikrosatelliten mit geeigneten Primerpaaren mittels PCR amplifiziert. Größenunterschiede, die aufgrund von Unterschieden in der

Länge des Motives entstehen, repräsentieren die unterschiedlichen Allele des Locus (Gupta et al. (1996)). Ihre Vorteile, wie die kodominante Vererbung, die hohe Redundanz, der hohe Umfang allelischer Diversität, die relativ leichte Abschätzung der Fragmentgrößen sowie die hohe Reproduzierbarkeit, haben diese Methode zu einer der beliebtesten molekularen Markertechnik werden lassen (Agarwal et al. (2008)). Ein großer Nachteil dieser Markertechnologie ist, dass locuspezifische Primerpaare für die meisten Arten neu entwickelt werden müssen (Zane et al. (2002)).

1.4 Entwicklung von Mikrosatellitenmarkern

Zu Beginn des Projektes im Jahr 2006 waren keine Mikrosatellitenmarker für *Phalaenopsis* veröffentlicht und/oder zugänglich. Innerhalb der letzten Jahre wurden jedoch mehrere Mikrosatellitenmarker entwickelt und veröffentlicht (Lin (2002) seit 2007 öffentlich zugänglich, Rodrigues and Kumar (2008)). Teilweise wurden diese Marker bereits in der vorliegenden Arbeit genutzt. Diese Mikrosatelliten wurden jedoch aus Wildarten entwickelt und wie im Laufe der vorliegenden Arbeit zu sehen sein wird, unterscheiden sich *Phalaenopsis* Wildarten und Hybriden so stark voneinander, dass die Anwendung dieser Mikrosatellitenmarker in *Phalaenopsis* Hybriden stark limitiert ist. Aus diesen Gründen sollten hier neue Mikrosatellitenmarker, die für die Analyse von *Phalaenopsis* Hybriden geeignet sind, entwickelt werden.

Squirrell et al. (2003) fassen die wichtigsten Schritte in der Entwicklung von Mikrosatellitenmarkern zusammen. Dabei werden die Erstellung einer Mikrosatelliten angereicherten DNA-Bank und die Testung der Primerpaare von den Autoren als die arbeits- und kostenintensivsten Schritte bezeichnet.

Von Zane et al. (2002) werden einige Methoden zur Erstellung einer nach Mikrosatelliten angereicherten DNA-Bank aufgeführt. Der Nutzen der Anreicherung wird durch die höhere Effizienz deutlich. Die Autoren zeigen, daß in Pflanzen eine durchschnittliche Effizienz von 2,3% (0,059-5,8%) Mikrosatellitenmotiv tragender Fragmente in einer nicht angereicherten DNA-Bank gefunden wird. Bei der Entwicklung von Mikrosatellitenmarkern aus einer angereicherten Bank von *Pelargonium* konnte beispielsweise eine Effizienz von 5-39% erreicht werden (Becher et al. (2000)).

In der vorliegenden Arbeit wurde eine, von Zane et al. (2002) als FIASCO (**F**ast **I**solation by **A**FLP of **S**equences **C**ontaining repeats) bezeichnete Anreicherungsstrategie angewandt. Bei dieser relativ schnellen und einfachen Methode werden DNA-Fragmente, ähnlich der von Vos et al. (1995) beschriebenen Methode, mit zwei Restriktionsenzymen geschnitten und mit Hilfe von an den Enden ligierten Adaptoren amplifiziert. Die Mikrosatellitenmotiv-tragenden Fragmente werden durch Hybridisierung

mit biotinylierten Sonden und anschließende Trennung über streptavidinummhüllte magnetische Kugeln (Magnetic Beads) selektiert. Die Fragmente werden ein weiteres Mal amplifiziert. Ein bis zwei weitere Anreicherungsrounds erhöhen die Effizienz.

Die Sequenzierung ausgewählter Klone wird heute bereits routinemäßig als Serviceleistung von verschiedenen Firmen übernommen (z.B. GATC Biotech AG, Konstanz). Bei der Konstruktion der Primerpaare sollten die üblichen Regeln bezüglich Länge, GC-Gehalt und der Vermeidung von Palindromen und repetitiven Primersequenzen beachtet werden. Des Weiteren sollten die Fragmente nicht länger als 100-250 bp sein, um ungenaue Amplifikationen zu vermeiden (Weising et al. (2005)).

Die Auftrennung von PCR-Produkten für die Mikrosatellitenanalyse erfolgt meist mit Hilfe eines Laser Detektion Systems. Dazu muss einer der Primer mit einer detektierbaren Markierung versehen sein, die meist relativ teuer ist (Agarwal et al. (2008)). Um diese Kosten zu verringern, wurde von Schuelke (2000) eine Methode entwickelt, bei der in einer Zwei-Schritt-PCR zunächst die Mikrosatellitenfragmente und diese anschließend mit Hilfe einer M13-Primerverlängerung und dem entsprechend markierten Primer ein zweites Mal amplifiziert werden. Aufgrund der M13-Markierung sind die Mikrosatellitenfragmente auf einem Laser Detektion System erfassbar. Es wird nur ein markierter Primer benötigt, der für alle zu testenden Primerpaare verwendet werden kann. Die von Schuelke entwickelte Technik wurde bereits in vielen Studien erfolgreich angewandt (z.B. Breseghello (2005), Wills and Burke (2007)).

1.5 Erstellung einer genetischen Karte

Das Ziel einer genetischen Karte ist es Gene oder Genregionen zu identifizieren, die einen Einfluss auf ein bestimmtes phänotypisches Merkmal haben, sowie die Beziehung dieser Gene bzw. Genregionen zueinander. Bei der physikalischen Kartierung werden die wahren Abstände in Basenpaaren (bp) auf einem Chromosom oder einer Region auf dem Chromosom ermittelt. Im Gegensatz dazu werden bei der genetischen Kartierung die Abstände aufgrund von Rekombinationsfrequenzen in centi Morgan (cM) festgelegt. Dabei ist die Wahrscheinlichkeit der Rekombination zwischen zwei Markern oder Genen umso größer, je weiter sie voneinander entfernt sind. Folglich zeigt die genetische Karte die relative Position von molekularen Markern bzw. Genen zueinander in so genannten Kopplungsgruppen (Linkage groups = LG). (Weising et al. (2005))

Da eine Kartierung, aufgrund der Anzahl an möglichen Allelen an einem Locus, einfacher in diploiden Organismen vorgenommen werden kann, werden Karten von Polyploiden meist in deren diploiden Verwandten erstellt (Wu et al. (1992)). Aller-

dings sind nicht immer diploide Organismen mit ausreichend hoher Verwandtschaft verfügbar. Wu et al. (1992) beschrieben erstmals die Verwendung von Single-Dose Restriction Fragments (SDRFs) zur Erstellung einer genetischen Karte in polyploiden Organismen. Neben RFLP-Markern werden auch andere Markertypen wie beispielsweise AFLP und SAMPL Marker (Barcaccia et al. (2003)) in der von Wu et al. (1992) beschriebenen Art und Weise zur Kartierung in polyploiden Organismen verwendet. Allgemein werden diese Marker heute als **Single Dose Marker** (SDM) bezeichnet. In der vorliegenden Arbeit werden neben AFLP- auch Mikrosatellitenmarker als SDM verwendet. Single Dose Marker werden aufgrund des Seggregationsverhältnisses in der Nachkommenschaft identifiziert. Ist ein Marker nur in einem der Eltern vorhanden, so spaltet dieser in der Nachkommenschaft in einem 1:1 Verhältnis. Ist der Marker jedoch in beiden Eltern vorhanden, so segregiert der Marker in einem 1:3 Verhältnis (Barcaccia et al. (2003)). Zusammenfassend ist ein SDM als DNA Fragment definiert, welches nur auf einem der homologen Chromosomen vorhanden ist und, unabhängig vom Ploidiegrad, in einem 1:1 Verhältnis (bzw. 1:3 Verhältnis bei biparentaler Vererbung) in der Nachkommenschaft aufspaltet (Stein et al. (2007)).

1.6 QTL-Analysen

Für die Ausprägung quantitativer Merkmale können mehrere Abschnitte eines Genoms (bzw. Gene) verantwortlich sein. Diese Abschnitte werden **Quantitative Trait Loci** (QTL) genannt. Da viele wichtige und interessante Eigenschaften bei Pflanzen quantitativ vererbt werden, liefern QTL-Analysen bzw. -Kartierungen wertvolle Informationen für die Verbesserung der Züchtung (Liu (1998)).

QTL-Analysen bzw. -kartierungen umfassen die Erstellung einer genetischen Karte sowie die Suche nach Beziehungen zwischen quantitativen Merkmalen und polymorphen Markern. Mit Hilfe der QTL-Analysen lassen sich Zusammenhänge zwischen quantitativen Eigenschaften und (polymorphen) Markermustern detektieren, ohne vorherige Kenntnisse über das/die entsprechende/n Gen/e zu haben. Faktoren, die einen Einfluss auf das entsprechende quantitative Merkmal haben, können entweder regulatorische Elemente und/oder auch Strukturgene sein. Neben Umwelteffekten bestimmen meist einige wenige 'major' Loci mit großem Einfluss und mehrere 'minor' Loci mit einem geringeren Einfluss die Ausprägung des quantitativen Merkmals.

Während für landwirtschaftliche Kulturen bereits viele QTL-Analysen durchgeführt wurden, wurden in Zierpflanzen bisher nur relativ wenige Studien durchgeführt (z.B. Linde et al. (2006), Dugo et al. (2005), Yan et al. (2007)).

Das Kopplungsungleichgewicht zwischen einem Marker und einem QTL dient als

Grundlage für die Detektion (Tanksley (1993)). Es werden hauptsächlich zwei Methoden zur Detektion von QTL angewandt. Eine Übersicht dieser beiden Methoden wird in Tanksley (1993) und Collard et al. (2005) gegeben. Bei der 'Einzel-Marker Analyse' werden einzelne Marker mit Hilfe statistischer Tests wie t-Test, Kruskal-Wallis, ANOVA und lineare Regression auf Zusammenhänge in der Vererbung mit der zu untersuchenden Eigenschaft getestet. Der Vorteil dieser Methode ist, dass keine genetische Karte notwendig ist. Allerdings ist es, aufgrund von Rekombination zwischen Marker und QTL, umso schwieriger einen QTL zu entdecken, je weiter der Marker von diesem entfernt ist. Bei der Intervallkartierung werden mehrere Marker gleichzeitig auf ihren Einfluss auf ein bestimmtes quantitatives Merkmal getestet. Da gekoppelte Marker gleichzeitig untersucht werden, können eventuelle Rekombinationen zwischen Marker und QTL in der Berechnung kompensiert werden. Für die Intervallkartierung ist eine genetische Karte nötig, in der die Anordnung der Marker möglichst gut abgesichert ist.

1.7 Assoziationsstudien

Zhu et al. (2008) definieren genomweite Assoziationsstudien als einen globalen Ansatz, um das Genom systematisch nach genetischen Variationen zu durchsuchen. In der einfachsten Form lässt sich die Assoziationsstudie folgendermaßen beschreiben: Eine Anzahl von Markern in einer Gruppe, die das zu untersuchende Merkmal trägt, wird mit einer Anzahl von Markern in einer anderen Gruppe, die dieses Merkmal nicht trägt, verglichen. Dabei sind die Individuen innerhalb sowie zwischen den Gruppen nicht miteinander verwandt. Schließlich wird auf Unterschiede in den Allelfrequenzen getestet. Man geht nun davon aus, daß Marker, die dicht an das Merkmal gekoppelt sind, Unterschiede in der Allelfrequenz zu den Markern zeigen, die nicht an das Merkmal gekoppelt sind. Die erfolgreiche Anwendung von Assoziationsstudien kann die Effizienz der Marker gestützten Selektion erhöhen, die Detektion von Genen vereinfachen und dabei helfen, Zusammenhänge in der Vererbung phänotypischer Merkmale und genetischer Sequenzen aufzuklären (Mackay and Powell (2007)).

Bei einer Assoziationsstudie oder auch Assoziationskartierung werden also Marker identifiziert, die einen signifikanten Unterschied in ihrer Allelfrequenz zwischen unverwandten Individuen zweier Gruppen aufweisen, wobei eine Gruppe das interessierende phänotypische Merkmal aufweist und die andere so genannte Kontrollgruppe nicht (Pritchard et al. (2000)). Ein entscheidender Vorteil in der Anwendung von Assoziationsstudien bei Pflanzen ist die Zeitersparnis, da keine Kartierungspopulation erstellt werden muss. Die Zeitersparnis wird besonders bei den Arten deutlich, deren

Reproduktionszeit relativ lang ist (Oraguzie et al. (2006)).

Um jedoch falsch positive Ergebnisse bei der Assoziationsstudie zu verringern / vermeiden, muss das Kopplungsungleichgewicht, welches durch die Populationsstruktur der zu untersuchenden Population entstanden ist, ermittelt werden (Zhu et al. (2008)). Dazu werden zunächst eventuell vorhandene Subpopulationen bestimmt und anschließend das Kopplungsungleichgewicht zwischen den Markern untersucht. Die Grundlagen dieser beiden Voraussetzungen werden in den folgenden Kapiteln kurz erläutert.

1.8 Diversitätsanalyse

In einer Diversitätsanalyse wird die genetische Diversität auf Grundlage der genetischen Distanz einer Population untersucht. Das Ausmaß der genetischen Diversität in einer Population kann von evolutionären Faktoren wie Mutation, Selektion, genetischer Drift und Migration aber auch von der Art der Vermehrung (Selbstbefruchter oder Fremdbefruchter) sowie von menschlichen Einflüssen, wie gezielter Züchtung, beeinflusst werden. Um das Potential der genetischen Ressource einer Population für die Züchtung entsprechend nutzen zu können, muss diese durch geeignete Methoden ermittelt werden.

Ursprünglich wurde die genetische Diversität anhand morphologischer Marker wie Blütenfarbe, Wuchshöhe oder Ähnliches ermittelt. Allerdings sind diese Marker stark umweltabhängig. Seit den 1960er Jahren besteht die Möglichkeit, die genetische Diversität mit Hilfe von Allo- und Isoenzymen zu untersuchen. Diese Marker haben den Vorteil, dass sie nicht von Umweltfaktoren beeinflusst werden. Sie sind jedoch nur in begrenzter Anzahl vorhanden und können geringe Variationen zwischen den Genotypen meist nicht detektieren. Mit der Entwicklung molekularer Marker war nun die Möglichkeit gegeben ausreichend Marker zur Verfügung zu haben, die nicht durch Umweltfaktoren beeinflusst sind und unabhängig vom Alter und Entwicklungsstadium der Pflanze erfasst werden können (Rao (2004)).

Die Eignung verschiedener molekularer Marker wurde bereits in mehreren Studien untersucht (z.B. Powell et al. (1996), Garcia-Mas et al. (2000), Belaj et al. (2003)). Mittlerweile werden AFLP und Mikrosatellitenmarker am häufigsten für die Analyse der genetischen Diversität genutzt. Um nun die Diversität zwischen den einzelnen Genotypen zu ermitteln, können drei Koeffizienten genutzt werden. Während der 'Simple Matching' Koeffizient auch die Abwesenheit von Banden in die Berechnung mit einbezieht, wird beim Jaccard (1908) und Dice (1945) bzw. Nei and Li (1979) Koeffizient nur die Anwesenheit der Banden berücksichtigt (Laurentin (2009)). Ver-

einfacht ausgedrückt werden mit Hilfe der Koeffizienten die paarweisen Distanzen zwischen den untersuchten Genotypen berechnet und als Distanzmatrix ausgegeben (Pritchard et al. (2000)). Zur Darstellung der Distanzmatrizen bzw. der individuellen Gruppen stehen verschiedene Methoden zur Verfügung. Eine der am meisten verwendeten Methoden ist die Neighbour-Joining Methode nach Saitou and Nei (1987). Hier werden die Gruppen auf Basis der jeweils kleinsten Distanzen gebildet. Um das konstruierte Dendrogramm statistisch abzusichern, hat sich eine Bootstrap-Analyse nach Felsenstein (1985) etabliert.

Inzwischen wurden bereits viele Analysen zur genetischen Diversität von fast allen wichtigen/interessanten Kulturen durchgeführt. Auch in Gattungen der Familie *Orchidaceae* sind eine Reihe von Diversitätsanalysen veröffentlicht worden (z.B. Xiang et al. (2003), Campbell et al. (2007), Wang et al. (2009), Verma et al. (2009)). Die genetische Diversität von *Phalaenopsis* sp. ist jedoch kaum untersucht. Bisherige Arbeiten untersuchten die genetische Diversität von *Phalaenopsis* Wildarten nur aufgrund von Teilen des Genoms (Tsai et al. (2003), Padolina (2006)). Erst vor kurzem wurde eine Studie bezüglich der Diversität von *Phalaenopsis* veröffentlicht, in der auch Hybriden und das Gesamtgenom betrachtet wurden (Chang et al. (2009)).

1.9 Populationsstruktur

In der Populationsstruktur werden die untersuchten Genotypen nicht nur aufgrund ihrer Distanzen zueinander in Zusammenhang gesetzt, es wird auch die Zugehörigkeit eines Genotyps zu einer bestimmten Population abgeschätzt. Dabei wird die Anzahl der Gruppen nicht nur visuell durch die Gruppenbildung im Dendrogramm ermittelt, sondern mit statistischen Methoden berechnet. Die Ermittlung der Populationsstruktur basiert auf folgendem Modell: Die Individuen innerhalb einer Population befinden sich im Hardy-Weinberg Gleichgewicht; daher die Marker aufgrund von Zufallspaarung im Kopplungsgleichgewicht. Werden nun Genotypen unterschiedlicher Populationen miteinander verglichen, so befinden sie sich zueinander nicht im Hardy-Weinberg-Gleichgewicht und die entsprechenden Allelfrequenzen sind zu Gunsten der jeweiligen Population verschoben. Dieses von Pritchard et al. (2000) aufgestellte Modell ist eine idealisierte Annahme, die im biologischen Sinn mit hoher Wahrscheinlichkeit in dieser Form nicht vorkommt. Im Unterschied zu der zuvor genannten Diversitätsanalyse werden die Genotypen hier auf Basis des Modells in Populationen bzw. Gruppen eingeteilt und nicht aufgrund von genetischen Distanzen.

Im einfachsten Fall ist die Populationsstruktur durch die geographische Herkunft oder durch phänotypische Unterschiede der Individuen zu ermitteln. Sind die Sub-

populationen jedoch nicht komplett voneinander getrennt und unabhängig, kann es durch Genfluss zu einer Durchmischung der genetischen Struktur kommen (Evanno et al. (2005)). Die Schwierigkeit besteht nun unter anderem auch darin so genannte kryptische Populationen zu ermitteln, welche nicht durch 'sichtbare' Eigenschaften ermittelt werden können, jedoch durchaus deutliche Unterschiede in der Genetik der Individuen verursachen können (Pritchard et al. (2000)).

In einer großen Population ohne Selektion, Mutation oder Migration, stellt sich das Hardy-Weinberg-Gleichgewicht nach einer Generation Zufallspaarung ein (Hardy (1908), Weinberg (1909)). Aufgrund von Flaschenhalseffekten bzw. genetischer Drift und Mutationen kann sich dieses Gleichgewicht jedoch erst später einstellen. Demzufolge hat eine Population, die solchen Einflüssen unterworfen war, einen höheren Anteil an Kopplungsungleichgewicht zwischen den Allelen, als eine Population, die diesen Effekten nicht unterworfen war (Nordborg and Weigel (2008)). Das heißt, je mehr sich Genotypen zueinander im Hardy-Weinberg-Gleichgewicht befinden, bzw. je geringer das Kopplungsungleichgewicht zwischen den entsprechenden Allelen ist, desto eher gehören diese Genotypen zu einer (Sub-) Population. Auf Grundlage dieser Annahmen werden in STRUCTURE 2.2 die Anzahl der Subpopulationen mit einem MCMC-Ansatz aus der Bayesianischen Statistik ermittelt (Pritchard et al. (2000)). Beim MCMC-Ansatz handelt es sich um eine 'Monte Carlo Markov Chain' auf deren Prinzip später genauer eingegangen wird (siehe Kapitel 2.5.6).

Das Programm STRUCTURE ist eines der am Meisten genutzten Programme zur Berechnung der Populationsstruktur. Dieses Programm basiert auf dem von Pritchard et al. (2000) postuliertem Modell und nach Erweiterungen von Falush et al. (2007) ist inzwischen Version 2.2 erhältlich.

Die hier beschriebene Methode zur Abschätzung der Populationsstruktur wurde bereits erfolgreich in *Triticum* (Maccaferri et al. (2005)), *Zea* (Reif et al. (2005)) und *Hordeum* (Presterl et al. (2007), Hübner et al. (2009)) angewandt. Es konnten Subpopulationen entsprechend ihrer Herkunft oder bestimmter Eigenschaften ermittelt werden. Ebenso wurde eine Übereinstimmung der Subpopulationen mit der Gruppenbildung in distanzbasierten Klusteranalysen gefunden.

1.10 Kopplungsungleichgewicht

Das Kopplungsungleichgewicht (Abkürzung 'LD' für Linkage Disequilibrium) ist definiert als 'nicht zufällige Assoziation von Allelen an separaten Loci', d.h. die Korrelation zwischen polymorphen Markern, die durch eine gleiche Historie von Mutation und Rekombination entstanden sind (Flint-Garcia et al. (2003)). Das Kopplungsun-

gleichgewicht kann aus der Populationsstruktur, Selektion, Drift oder physikalischer Kopplung entstehen bzw. von diesen beeinflusst werden. Das Ausmaß des Kopplungsungleichgewichts hat dabei einen entscheidenden Einfluss auf die mögliche Auflösung und den notwendigen Aufwand einer Assoziationsstudie (Oraguzie et al. (2006)).

Die Berechnung des Kopplungsungleichgewichts kann nach der Formel von Lewontin (1964) erfolgen:

$$D' = \begin{cases} \frac{D}{\min(p_1 q_2, p_2 q_1)} & D > 0 \\ \frac{D}{\min(p_1 q_1, p_2 q_2)} & D < 0 \end{cases} \quad (1.1)$$

Häufiger wird das Kopplungsungleichgewicht jedoch als r nach der Formel von Hill and Weir (1994) angegeben:

$$r = \frac{D}{(p_1 p_2 q_1 q_2)^{1/2}} \quad (1.2)$$

D beschreibt die Differenz zwischen der Gametenfrequenz in Kopplung und der Gametenfrequenz in Repulsion. Die jeweilige Allelfrequenz eines Locus ist mit p und q dargestellt (Oraguzie et al. (2006)).

Das Kopplungsungleichgewicht einer Gattung hängt von vielen Faktoren, unter anderem von der Rekombinationswahrscheinlichkeit, ab. Die Graphik in Abbildung 1.2 zeigt den Abfall des Kopplungsungleichgewichts in Abhängigkeit von der Zeit bzw. der Anzahl an Generationen. Dabei stellt θ den Rekombinationsanteil zwischen zwei Loci dar. Sind zwei Loci voneinander unabhängig, also liegen sie z.B. auf zwei verschiedenen Chromosomen, so beträgt $\theta=0.5$ und es braucht nur wenige Generationen bis sich die beiden Loci im Gleichgewicht befinden. Sind zwei Loci hingegen gekoppelt, also liegen sie dicht beieinander, so ist θ (die Rekombinationswahrscheinlichkeit) klein. Diese Loci brauchen also wesentlich mehr Generationen / Rekombinationsereignisse, bevor sie sich im Gleichgewicht befinden. Je kleiner also θ ist, desto länger brauchen die Loci um ins Gleichgewicht zu kommen. Ist eine Mutation gerade erst geschehen (Generation = 0) so befindet sich der mutierte Locus mit allen anderen Loci im Ungleichgewicht ($D=1$). Je dichter also ein Locus an einer Mutation liegt, bzw. je dichter ein Marker an dem Locus, der für die untersuchte Eigenschaft verantwortlich ist, liegt, desto mehr befinden sich die beiden Loci im LD. Wenn es jetzt jedoch zu einem Flaschenhalseffekt bzw. gezielter Züchtung kommt, ist die Paarung nicht mehr zufällig sondern auf die, dem Flaschenhalseffekt ausgesetzte Population, begrenzt. Damit ist die Reduktion des LD begrenzt bzw. verlangsamt worden. Aus

diesem Grund muss die Populationsstruktur der zu untersuchenden Population ermittelt werden, um diese mit in die Berechnungen mit einfließen zu lassen. Weitere, das Kopplungsungleichgewicht beeinflussende Faktoren sind in Tabelle 4.1 aufgeführt und werden dort diskutiert.

Das Kopplungsungleichgewicht und deren Anwendung in Assoziationsstudien wurde in mehreren Veröffentlichungen ausführlich beschrieben (z.B. Sorkheh et al. (2008); Jorde (2000); Gupta et al. (2005); Flint-Garcia et al. (2003); Mackay and Powell (2007); Oraguzie et al. (2006)).

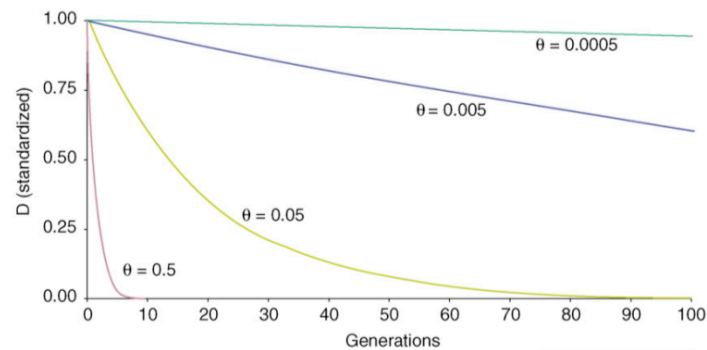


Abbildung 1.2: **Abfall des Kopplungsungleichgewichts in Abhängigkeit von der Zeit/Anzahl der Generationen** (aus: Mackay and Powell (2007)); D =Koeffizient zur Messung des Kopplungsungleichgewichts; θ =Rekombinationsfrequenz; je höher die Rekombinationsfrequenz, desto schneller reduziert sich das Kopplungsungleichgewicht zwischen zwei Loci

1.11 Ziel der Arbeit

Ziel dieser Arbeit ist es, grundlegende Kenntnisse über die genomische Struktur der heute auf dem Markt befindlichen *Phalaenopsis* Hybriden zu erhalten. Zur Verbesserung der Züchtung sollen züchterisch wichtige Eigenschaften, wie z.B. Blütengröße oder Rispenlänge mit molekularen Markern assoziiert und kartiert werden. Weitere Ziele sind die Aufklärung der Populationsstruktur und der Diversität zwischen den Hybriden.

2 Material und Methoden

2.1 Pflanzenmaterial

Zur Entwicklung von Mikrosatellitenmarkern wurde der *Phalaenopsis*-Genotyp 0744 (Abb. 2.1 a) als Topfpflanze von der Firma Bock Bio Science GmbH zur Verfügung gestellt. Mehrere Pflanzen dieses Genotyps wurden zusammen mit der Nachkommenschaft '21x19' (Abb. 2.1 b) unter den dort genannten Gewächshausbedingungen kultiviert.

Eine Nachkommenschaft von zwei frei auf dem Markt erhältlichen Genotypen wurde von Prof. Dr. Traud Winkelmann (Institut für Zierpflanzen- und Gehölzwissenschaften der Leibniz Universität Hannover) zur Verfügung gestellt. Diese Nachkommenschaft wurde durch eine Kreuzung zwischen dem Genotyp 19 (Blütenfarbe: weiß) und Genotyp 21 (Blütenfarbe: violett) erzeugt. Die Pflanzen wurden zunächst von *in vitro* Kultur in ein handelsübliches Vermehrungssubstrat mit einer Aufdüngung von 0,8 g/l Mehrnährstoffdünger überführt. Nach ca. 22 Wochen wurden die Pflanzen in Einzeltöpfe in Torfkultursubstrat I ebenfalls mit einer Aufdüngung von 0,8 g/l getopft. Nach drei Wochen wurde mit der Flüssigdüngung (dreimal wöchentlich) mit Universol blau (18:11:18:25) 0,1%ig begonnen. Die Temperatur rangierte zwischen 22 und 25 °C und wurde, zur Blüteninduktion, für acht Wochen auf 19/17 °C Tag/Nacht Temperatur abgesenkt. Anschließend wurde die Temperatur wieder leicht (21 °C) angehoben. Für die in dieser Arbeit vorgestellten Analysen wurden 103 Nachkommen der Kreuzung '21x19' verwendet.

Für die Analyse der Populationsstruktur und der Assoziationskartierung wurden 134 *Phalaenopsis* Hybriden und 10 Wildarten ebenfalls von der Firma Bock Bio Science GmbH zur Verfügung gestellt (siehe Tabelle 5.15 bis 5.23). Es handelte sich hierbei um eine Sammlung von heute auf dem Markt befindlichen Hybriden. Zur DNA Isolierung wurden Pflanzen aus *in vitro* Kultur verwendet. Von den Wildarten wurden Blattsegmente von im Gewächshaus der Firma Bock Bio Science GmbH kultivierten Pflanzen genutzt.



(a) Genotyp 0744



(b) Nachkommenschaft

Abbildung 2.1: (a) Blüte des Genotyps '0744': Dieser Genotyp wurde für die Entwicklung von Mikrosatellitenmarkern verwendet. (b) Blüten der Genotypen '19' und '21' und deren Nachkommen der Kreuzung '19x21' (links) sowie '21x19' (rechts). Für die in dieser Arbeit vorgestellte Analyse wurde die Kreuzung '21x19' verwendet.

2.2 Sonstige Materialien

Eine Aufstellung der verwendeten Geräte, Chemikalien und Enzyme sowie die Zusammensetzung der verwendeten Puffer und Lösungen ist dem Anhang zu entnehmen.

2.3 Aufnahme der phänotypischen Daten

Die phänotypischen Daten der Genotypensammlung wurden von der Firma Bock Bio Science GmbH zur Verfügung gestellt (siehe Tabelle 2.1).

Von der Nachkommenschaft '21x19' wurden die phänotypischen Daten von den im Gewächshaus kultivierten Pflanzen aufgenommen. Die Blütengröße und die Rispenlänge wurden jeweils in Millimeter (mm) gemessen. Dabei wurde der horizontale Durchmesser aller an der Rispe befindlichen Blüten gemessen und der Mittelwert gebildet. Auch bei der Rispenlänge wurde der Mittelwert gebildet, sofern sich mehr als eine Rispe an der Pflanze befand. Als weiteres Merkmal wurde der Anthocyangehalt der Blüten (Tepalen und Lippen) und der Blätter erfasst.

Für die Bestimmung des Anthocyangehaltes wurden vom jüngsten Blatt jeweils an der Spitze, an der Mittelrippe und am Blattrand eine Blattscheibe von 0,5 cm Durchmesser mit Hilfe eines Korkbohrers entnommen. Von der Blüte wurde jeweils eine Scheibe von 1 cm Durchmesser von drei Tepalen sowie mit einem Durchmesser von 0,8 cm von der Lippe entnommen. Wenn möglich, wurden mehrere Blüten beprobt und der Mittelwert der gemessenen Werte berechnet. Die Blattscheiben wurden einzeln in

Tabelle 2.1: Aufnahme und Ausprägung der phänotypischen Merkmale in der Genotypensammlung (erhoben von Bock Bio Science GmbH)

Merkmal	Daten	Ausprägung
Herkunft	ordinal	15 verschiedene Züchter (A-R kodiert)
Blütenfarbe	ordinal	weiß; weiß mit grün/gelb; weiß mit rot/violett; gelb; gelb mit rot/violett; pink; rosa; apricot/lachs; lila/violett; orange; bordeaux/rot
Blütentyp	ordinal	amabilisförmig; geschlossen; offen; sternförmig; Strahlen; stuartianaförmig
Blütengröße	nominal	3-12 cm
Blütenkonsistenz	ordinal	normal; fest; fleischig
Blattform	ordinal	länglich; rund
Blatttyp	ordinal	aufrechte; hängende Blattstellung
Blattzeichnung	ordinal	ja; nein
Anthocyane im Blatt	ordinal	ja; nein
Rispenlänge	nominal	10-75 cm
Duft	ordinal	ja; nein

jeweils 1 ml Methanol:Salzsäure (99:1) für 30 min bei Raumtemperatur und konstanter Bewegung inkubiert. Anschließend wurden die Blattscheiben aus der Lösung entfernt. Der Anthocyangehalt wurde indirekt durch die Absorption bei 525 nm im Photometer gemessen und wird im folgenden in OD (Optische Dichte) angegeben.

2.4 Molekularbiologische Methoden

2.4.1 DNA-Extraktion

Die Extraktion der DNA erfolgte nach der von Doyle and Doyle (1987) erstmals publizierten CTAB-Methode, die für die Anwendung bei *Phalaenopsis* modifiziert wurde.

Je nach Größe der Pflanze wurde bis zu 2 g möglichst junges Blattmaterial in flüssigem Stickstoff gefroren und mit Hilfe von zuvor ebenfalls in Stickstoff gekühlten Stahlkugeln und -gefäßen 3 min in einer Schwingmühle bei 25-30 Schwingungen pro Sekunde fein zermahlen. Das Pulver wurde in noch gefrorenem Zustand mit auf 65°C vorgeheiztem CTAB-1 Puffer (10 ml Puffer pro 1 g Blattmaterial) vermischt und anschließend im Wasserbad bei 65°C für 30 min inkubiert. Die Suspension wurde mit 1 Vol. Chloroform gemischt und für 20 min bei 4500 Upm zentrifugiert. Die obere wässrige Phase wurde abgenommen, mit 0,5 Vol. Chloroform gemischt und für 20 min bei 4500 Upm zentrifugiert. Dieser Schritt wurde einmal wiederholt. Die DNA wurde im folgenden Schritt mit 0,6 Vol. Isopropanol über Nacht bei -20°C

gefällt und anschließend 5 min bei 5 000 Upm abzentrifugiert. Das entstandene Pellet wurde mit 70% Ethanol (unvergällt) gewaschen und in 2 ml TE-Puffer (1x) gelöst. Durch Zugabe von 3 μ l RNase A (10 mg/ml) und einer Inkubation von 30 min bei 37°C wurde noch vorhandene RNA verdaut. Um verbliebene Proteine aus der DNA-Lösung zu entfernen, wurde 1 Vol. Phenol/Chloroform (1:1) dazugegeben und nach gründlicher Durchmischung bei 13 000 Upm für 5 min zentrifugiert. Dieser Schritt wurde mit Chloroform wiederholt, um verbliebene Phenolreste zu entfernen. Durch Zugabe von 1/10 Vol. 3 M NaAc und 2 Vol. 98% Ethanol (unvergällt), Durchmischung und anschließender Zentrifugation bei 5 000 Upm wurde die DNA erneut gefällt. Nach einem erneuten Waschschrift mit 70% Ethanol (unvergällt) wurde die DNA in 100 μ l TE-Puffer (1x) pro 1 g eingesetzten Pflanzenmaterial aufgenommen.

Quantifizierung von Nukleinsäuren

Die Quantifizierung der isolierten DNA erfolgte mittels Agarose Gelelektrophorese und der Software Gel-Pro Analyzer (Media Cybernetics). Als Mengenstandard wurde λ -DNA in den Mengen 10, 20, 40, 60, 80 und 100 ng verwendet. Um eine möglichst genaue Quantifizierung zu erhalten, wurden neben der Originallösung verschiedene Verdünnungsstufen (1:2, 1:5, 1:10) der isolierten DNA mit dem Standard verglichen und die Konzentration als Mittelwert über die verschiedenen Verdünnungsstufen berechnet.

2.4.2 Polyacrylamid Gelelektrophorese

Zur Auftrennung und Detektion von Infrarot Farbstoff (IRD) markierten DNA Fragmenten (AFLP und Mikrosatelliten) diente die Polyacrylamid Gelelektrophorese, welche entweder mit dem 4200 oder 4300 DNA Analyzer System von LiCor® Bioscience durchgeführt wurde. Als Matrix diente ein 6% denaturierendes Polyacrylamidgel (SequaGel™ XR, Biozym Scientific GmbH). Je nach Größe und Anzahl bzw. Dichte der Fragmente wurde eine Laufstrecke zwischen 15 und 30 cm sowie eine Dicke von 0,25 mm gewählt. Entsprechend der zu erwartenden Signalstärke der Fragmente wurden die DNA Proben mit der entsprechenden Menge Ladepuffer (siehe Anhang) versetzt. Die Auftrennung erfolgte innerhalb von 30 min bis zu 7 Stunden (je nach Größe der Fragmente und Länge der Laufstrecke) bei 1500 V, 35 mA, 35 W und 45 °C. Die Dokumentation der Fragmente erfolgte automatisch mit der e-Seq V2.0, e-Seq V3.0 oder SAGA Generation 2 Software. Es folgte eine manuelle Auswertung der Daten.

2.4.3 Herstellung einer für Mikrosatelliten angereicherten DNA Bank

Für die Herstellung einer für Mikrosatelliten angereicherten DNA Bank wurde der Genotyp 0744 verwendet. Die DNA wurde wie oben beschrieben isoliert und quantifiziert. Die Anreicherung der Mikrosatelliten-tragenden Fragmente wurde nach der Methode von Guo-Qing et al. (2003) mit Modifikationen durchgeführt. Zunächst wurde die DNA äquivalent zur AFLP-Technik vorbereitet. Wie in Kapitel 2.4.5 beschrieben, wurden Restriktion, Ligation und Präamplifikation durchgeführt. Als einzige Modifikation wurde der Ansatz für die Präamplifikation verdoppelt, um ausreichend Fragmente für die folgenden Schritte zur Verfügung zu haben.

Anreicherung von Mikrosatelliten-Motiv-tragenden Fragmenten

Der erste Schritt der Anreicherung ist die Hybridisierung von biotinylierten Sonden an die präamplifizierte DNA-Fragmente. Dazu wurden Oligonucleotide mit den Motiven (GT)₁₀ und (AG)₁₀ und einer Biotinmarkierung verwendet. Für die Hybridisierung wurden die in Tabelle 2.2 aufgezeigten Ansätze in einem Gesamtvolumen von 22 µl verwendet. Da nicht bekannt war, welche Menge an präamplifizierter DNA zu einem guten Ergebnis führt, wurden zwei Ansätze mit unterschiedlichem DNA-Puffer-Verhältnis getestet.

Tabelle 2.2: Eingesetzte Mengen für die Hybridisierung der biotinylierten Sonden an die DNA-Fragmente

	Ansatz 1	Ansatz 2
präamplifizierte DNA	2 µl	10 µl
biotinylierte Sonde	2,5 µM	2,5 µM
6x SSC + 0,1% SDS	17,5 µl	9,5 µl

Bis auf das unterschiedliche DNA-Puffer-Verhältnis wurden beide Ansätze in den folgenden Schritten gleich behandelt. Zunächst wurde die DNA im Reaktionsgemisch (siehe Tabelle 2.2) bei 95 °C für 5 min denaturiert. Um ein Renaturieren der DNA Fragmente zu verhindern, wurde das Gemisch sofort für 3-5 min auf Eis abgekühlt. Die Anlagerung der biotinylierten Sonden erfolgte durch 60 min Inkubation bei 60 °C.

Die Selektion der Mikrosatelliten tragenden Fragmente erfolgte nun indirekt mit Hilfe von 'Magnetic Beads' (Dynabeads M-280). Zunächst wurden 20 µl (10 µg/µl) 'Magnetic Beads' viermal mit 500 µl 6x SSC gewaschen und anschließend in 100 µl 6x SSC gelöst. Die DNA mit den hybridisierten Sonden wurde vollständig zu den gelösten 'Magnetic Beads' gegeben und alles gut durchmischt. Es folgte eine Inkubation von 30 min bei Raumtemperatur, bei der die Lösung zwischenzeitlich durch

Anschnipsen durchmischt wurde. Da die 'Magnetic Beads' mit Streptavidin umhüllt sind und Streptavidin eine hohe Affinität zu Biotin hat, binden die Sonden und somit die Fragmente, an denen sie hybridisiert sind, an den 'Magnetic Beads'. Mit Hilfe eines Magneten wurden die 'Magnetic Beads' pelletiert. Alle nicht gebundenen Fragmente wurden durch die in Tabelle 2.3 aufgeführten Waschschr tte entfernt:

Tabelle 2.3: **Waschschr tte zum Entfernen nicht gebundener Fragmente:** um die nicht gebundenen Fragmente zu entfernen werden nacheinander die hier aufgef hrten L sungen verwendet

Anzahl	Menge	L�sung	Temperatur	Zeit
1x	400 μ l	6x SSC + 0,1% SDS	60 $^{\circ}$ C	15 min
2x	400 μ l	6x SSC + 0,1% SDS	Raumtemperatur	15 min
1x	500 μ l	2x SSC	Raumtemperatur	~5 min
1x	500 μ l	1x SSC	Raumtemperatur	~5 min

Es folgte eine Elution der Fragmente bei 5 min und 90 $^{\circ}$ C in 40 μ l bidest Wasser. Die Magnetic Beads wurden mit Hilfe des Magneten pelletiert und der  berstand, mit den angereicherten Fragmenten, in ein neues Gef   berf hrt. Von dieser L sung wurden 5 μ l f r die zweite Pr amplifikation verwendet.

Die Anreicherung und anschließende Pr amplifikation wurden ein- bis zweimal wiederholt (Ergebnisse siehe Tabelle 3.2 zweite bzw. dritte Anreicherungsrunde). Nach jeder Pr amplifikation wurden 10 μ l des Pr amplifikats auf einem 1%igem Agarosegel aufgetrennt. Aufgrund unterschiedlicher Fragmentgr  en war das Pr amplifikat als Schmier sichtbar. Die dritte bzw. vierte Pr amplifikation wurde vollst ndig (50 μ l) auf einem 1%igem Agarosegel aufgetrennt und nach einem Gr  envergleich mit einer 1 kb-Leiter wurden alle Fragmente gr  er als 500 bp ausgeschnitten und mit Hilfe des GeneClean[ ] Kit von Q Biogene nach beiliegendem Protokoll aufgereinigt.

Klonierung von DNA-Fragmenten mit Mikrosatelliten

Als Klonierungsvektor wurde der pGEM[ ]-T Easy Vektor von Promega verwendet. Die, wie in Kapitel 2.4.3 beschrieben, angereicherten und aufgereinigten Fragmente wurden nach Herstellerprotokoll in den pGEM[ ]-T Easy Vektor inseriert. Das Gesamtvolumen des Ligationsansatzes wurde dabei auf 5 μ l reduziert und ein Insert:Vektor-Verh ltnis von 3:1 verwendet. Die Ligation erfolgte  ber Nacht bei 4 $^{\circ}$ C.

Der f r die anschließende Transformation verwendete *E. coli* Bakterienstamm ist im Anhang in Tabelle 5.1 aufgef hrt.

F r die Transformation wurden 1 μ l Ligationsansatz mit 40 μ l der elektrokompenten Zellen vermischt. Das Gemisch wurde in Elektroporationsk vetten mit 2 mm Elektrodenabstand  berf hrt. Die Transformation erfolgte in einem *E. coli*-Pulser

(Biorad) bei 2,5 kV. Nach der Elektroporation wurde der Ansatz sofort mit 1 ml SOC-Medium gemischt und es folgte eine Erholungsphase von 1 Stunde bei 37 °C und 150 Upm in 15 ml Falcongefäßen. Anschließend wurden 50-200 µl der Bakteriensuspension in Petrischalen mit festem LB-Medium ausgestrichen. Für die Selektion wurde dem LB-Medium zusätzlich 50 µg/ml Ampicillin, 16 µM IPTG, sowie 32 mM X-Gal beigemischt. Die transformierten Bakterien wurden über Nacht bei 37 °C inkubiert. Für eine deutlichere Ausfärbung der Kolonien wurden die Petrischalen für 1-2 Stunden bei 4 °C gelagert.

Mittels Blau-Weiß-Selektion wurden Kolonien selektiert, welche ein Plasmid mit DNA Insert enthalten. Diese Bakterienkolonien wurden mit einem sterilen Zahnstocher gepickt und in 70 µl flüssiges, ampicillinhaltiges FB-Medium in 384er Mikrotiterplatten überführt. Nach einer erneuten Inkubation über Nacht bei 37 °C wurden die Kolonien bei -80 °C gelagert.

Identifizierung der Mikrosatelliten-Motiv tragenden Fragmente

Zur Identifizierung von klonierten DNA Fragmenten, die ein Mikrosatelliten-Motiv tragen, wurden mehrere Teilschritte durchgeführt, die im Folgenden kurz erläutert werden.

Kolonie-Blot

Mit Hilfe eines sterilen 384er Replikationsstempels wurde die Bakteriensuspension auf eine zuvor zurecht geschnittene Nylonmembran (Hybond-N, Amersham) übertragen. Diese Membran wurde auf festem, Ampicillin-haltigem (50 µg/ml) LB-Medium bei 37 °C über Nacht inkubiert. Zur Freisetzung, Denaturierung und Fixierung der Plasmid DNA wurde die Membran nacheinander für jeweils 5 min bei Raumtemperatur mit in folgenden Lösungen getränktem Whatman-Papier inkubiert: einmal in 10% SDS, zweimal in Denaturierungslösung und einmal in Neutralisierungslösung (Rezepte siehe Anhang). Anschließend wurden die Membranen für 5 min in 2x SSC gewaschen. Die Fixierung der DNA erfolgte bei 80 °C für 2 Stunden.

Herstellung der Sonden

Aufgrund der Anreicherung der Fragmente mit (GT)₁₀ und (AG)₁₀ Oligonukleotiden wurde nach Fragmenten gesucht, die diese Motive enthalten. Um eine möglichst hohe Anzahl an Wiederholungen der Mikrosatellitenmotive zu finden, wurden die Oligonukleotide (GT)₁₅ und (AG)₁₅ verwendet.

In einem Gesamtvolumen von 25 µl wurden 0,4 µM Oligonukleotide, 1/5 Vol. 5x forward Puffer (BRL), 1,85 MBq [γ -³²P]-dATP (Hartmann Analytic) und 10 U T4-Kinase gemischt und anschließend für 15 min bei 37 °C inkubiert. Um die Reaktion zu stoppen bzw. nicht eingebaute Nukleotide zu entfernen, wurde dem Gemisch

1 μ l 0,5 M EDTA (pH 8,0) und 80 μ l 1x TE zugefügt und durchmischt. Anschließend wurde das Gemisch für 2 min bei 3000 Upm über eine Sephadex G50-Säule zentrifugiert. Die ungefähre Einbaurate der radioaktiv markierten Nukleotide wurde mit einem Handmessgerät für radioaktive Strahlung gemessen.

Hybridisierung

Zunächst wurden die Bakterien mit einer Lösung aus 1x SSC und 0,1% SDS vom Filter gewaschen. Die Filter wurden anschließend in feuchtem Zustand in Hybridisierungsröhren gelegt. Zur Absättigung unspezifischer Bindungsstellen auf den Filtern wurden diese mit 20 ml Hybridisierungspuffer für 1 Stunde bei 55 °C und konstanter Rotation inkubiert. Für die Hybridisierung mit den radioaktiv markierten Sonden wurden diese zum Hybridisierungspuffer in die Hybridisierungsröhre pipettiert. Es folgte eine Inkubation über Nacht bei 55 °C und konstanter Rotation.

Am nächsten Tag wurden die Membrane mit einer Lösung aus 6x SSC und 0,1% SDS dreimal für jeweils 30 min bei 25 °C und einmal für 5 min bei 45 °C gewaschen.

Die Exposition der Filter erfolgte in Kodak X-Omatic-Kassetten auf HyperfilmTM MP Autoradiographie-Filmen bei -80 °C für ein bis mehrere Tage, je nach Signalstärke. Die Filme wurde anschließend mit Entwickler und Fixierer von Kodak entwickelt. Nur starke Signale wurden als positiv gewertet, da schwache Signale aufgrund einer zu geringen Stringenz beim Waschen entstanden sein können.

Insert PCR

Positiv bewertete Klone wurden selektiert und die Größe des Inserts wurde durch PCR-Amplifikation überprüft. Dazu wurden M13-Primer verwendet, die jeweils links und rechts neben der Klonierungsstelle des Vektors binden. Die PCR wurde in einem Reaktionsansatz mit einem Gesamtvolumen von 25 μ l und folgenden Komponenten durchgeführt: jeweils 0,25 μ M vorwärts und rückwärts M13-Primer, 0,2 mM dNTPs, 0,5 U Taq-Polymerase (Firepol) und 1/10 Vol. 10x Williams Puffer. Eine Überführung der zu testenden Bakterien erfolgte direkt von der Bakteriensuspension in das Reaktionsgemisch mit Hilfe eines sterilen Zahnstochers oder eines sterilen 96er Replikationsstempels. Die Reaktionsbedingungen sind Tabelle 2.4 zu entnehmen.

Tabelle 2.4: Reaktionsbedingungen Insert PCR

Anzahl der Zyklen	Reaktionsschritt	Zeit	Temperatur
1	Anfangsdenaturierung	2 min	95 °C
30	Denaturierung	45 sek	94 °C
	Hybridisierung	45 sek	58 °C
	Synthese	2 min	72 °C
1	Endsynthese	10 min	72 °C

Zur Größenbestimmung der amplifizierten Fragmente wurde der PCR Gesamtansatz sowie eine 1 kb Leiter als Größenstandard auf einem 1%igen Agarosegel aufgetrennt. Da neben dem integrierten DNA Fragment auch ein kleines Stück vom Vektor mit amplifiziert wird, wurde dies bei der Bestimmung der Größe der Fragmente berücksichtigt.

Sequenzierung von DNA-Fragmenten

Für die Sequenzierung der DNA-Fragmente mit Mikrosatelliten wurden ausschließlich Kolonien selektiert, welche bei der Insert-PCR eine Fragmentgröße von mindestens 500 bp aufwiesen. Diese Bakterien wurden auf Ampicillin-haltigen LB-Festmedium in 96er Mikrotiterplatten übertragen und über Nacht bei 37 °C inkubiert. Die Sequenzierung der integrierten Fragmente wurde von der Firma GATC Biotech AG durchgeführt. Die weitere Verwendung der Sequenzen wird im Abschnitt Bioinformatik (2.5) erläutert.

2.4.4 Testen und Anwendung der Mikrosatelliten-Primerpaare

Aus den in Abschnitt 2.4.3 beschriebenen Sequenzen wurden, wie in Abschnitt 2.5.1 und 2.5.2 beschrieben, Primerpaare entwickelt, die auf ihre Funktionalität zu überprüfen waren. Hierzu wurde die Methode nach Schuelke (2000) angewandt. Abbildung 2.2 zeigt die schematische Darstellung dieser Methode. Bei dieser Methode wird an das 5'-Ende des vorwärts Primers zusätzlich folgende Nukleotidsequenz angehängt: GTA AAA CGA CGG CCA GT. Diese Sequenz dient als Anlagerungsstelle für den vorwärts M13-Primer, welcher mit einer IRD Markierung versehen wurde. Aufgrund dieser Markierung und dem Einbau des M13-Primers in die amplifizierten Fragmente (siehe Abbildung 2.2) konnten diese in einer Polyacrylamid Gelelektrophorese auf einem LiCor Sequenz System detektiert werden.

Die zu testenden Primerpaare wurden mit der entsprechenden Verlängerung von der Firma Eurofins MWG Operon synthetisiert. Das Reaktionsgemisch mit einem Gesamtvolumen von 15 µl besteht aus folgenden Komponenten: 10-20 ng DNA, 1/5 Vol. 10x Williams Puffer, 0,08 mM dNTPs, 0,004 µM vorwärts Primer mit M13-Verlängerung, 0,02 µM rückwärts Primer, 0,067 µM vorwärts M13-Primer, 0,6 U Taq-Polymerase (Firepol oder Boline). In Tabelle 2.5 sind die Reaktionsparameter aufgeführt.

Zum Testen der Primerpaare wurden die Genotypen 0744, 1307, 19 und 21 verwendet. Zunächst wurde eine Gradienten-PCR mit den in Tabelle 2.5 aufgeführten Parametern durchgeführt, um die optimale Hybridisierungstemperatur zu ermitteln. Anschließend wurde die entsprechende Zyklenanzahl sowie Hybridisierungstemperatur für die entsprechenden Primerpaare festgelegt und für die in dieser Arbeit be-

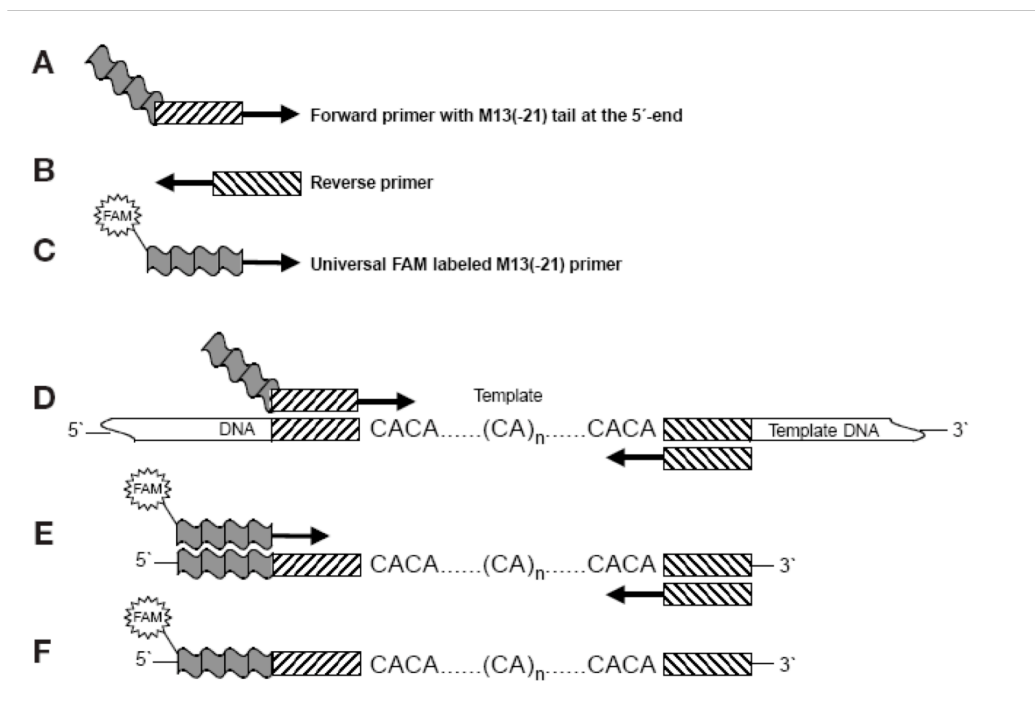


Abbildung 2.2: **Amplifikationsschema der M13-Mikrosatelliten PCR** (aus Schuelke (2000)): (A) vorwärts Primer mit M13-Verlängerung am 5'-Ende, (B) rückwärts Primer, (C) M13-Primer mit IRD-Markierung, (D) erste Amplifikation: Anlagerung der Primer an die flankierenden Sequenzen des Mikrosatellitenmotivs, (E) zweite Amplifikation: Anlagerung des M13-vorwärts-Primers sowie des rückwärts Mikrosatellitenprimers an die zuvor amplifizierten Fragmente, (F) IRD-markiertes Mikrosatellitenfragment

Tabelle 2.5: Reaktionsbedingungen für die M13-Mikrosatelliten PCR

Anzahl der Zyklen	Reaktionsschritt	Zeit	Temperatur
1	Anfangsdenaturierung	5 min	94 °C
20-25	Denaturierung	45 sek	94 °C
	Hybridisierung	1 min	56-63 °C
	Synthese	1 min	72 °C
8	Denaturierung	30 sek	94 °C
	Hybridisierung	45 sek	52 °C
	Synthese	1 min	72 °C
1	Endsynthese	10 min	72 °C

schriebenen populationsgenetischen Untersuchungen angewandt.

Um den Informationsgehalt der Mikrosatellitenmarker bewerten zu können, wird der so genannte 'Polymorphism Information Content' (PIC) für verschiedene Genotypen nach Botstein et al. (1980) mit folgender Formel berechnet:

$$PIC = 1 - \sum_{i=1}^n p_i^2 - \sum_{i=1}^{n-1} \sum_{j=i+1}^n 2p_i^2 p_j^2 \quad (2.1)$$

Der Grad der Heterozygotie (H) unterscheidet sich nur marginal vom PIC-Wert und wird mit folgender Formel berechnet:

$$H = 1 - \sum_{i=0}^n p_i^2 \quad (2.2)$$

Dabei entspricht p_i der Frequenz des i-ten Allels, p_j der Frequenz den j-ten Allels und n ist die Anzahl der Fragmente.

2.4.5 Anpassung und Anwendung der AFLP-Technik

Das in dieser Arbeit angewendete AFLP Protokoll basiert auf der von Vos et al. (1995) entwickelten Methode. Es wurde auf die im Labor üblichen Praxis und Gegebenheiten sowie auf das *Phalaenopsis*-Genom modifiziert.

Grundlegend besteht die AFLP-Technik aus vier Teilschritten, die im Folgenden kurz erläutert werden. Zunächst wurde die *Phalaenopsis* DNA mit den Restriktionsenzymen *Hind*III und *Mse*I verdaut. Dazu wurden 200-250 ng DNA mit 5 U *Hind*III, 3 U *Mse*I und 1/10 Vol. 10x RL-Puffer in einem Gesamtvolumen von 25 μ l über Nacht bei 37 °C verdaut. Es folgte eine Ligation von *Hind*III- und *Mse*I-Adapter (siehe Anhang 5.3). Ein Ligationsgemisch (0,5 μ M *Hind*III-Adapter, 5 μ M *Mse*I-Adapter, 1/10 Vol. 10x RL-Puffer, 0,25 U T4-DNA-Ligase) mit dem Gesamtvolumen von 5 μ l wurde dem Restriktionsgemisch zugegeben, vermischt und für 4,5 Stunden bei 37 °C inkubiert. Dabei ligierten die Adapterkassetten an die entsprechenden Enden der DNA-Fragmente und die Sequenz der Adapter diente in den folgenden Amplifikationen als Primeranlagerungssequenz.

Für die Präamplifikation wurden 5 μ l der verdauten und ligierten DNA in einem Gesamtvolumen von 50 μ l eingesetzt. Des Weiteren enthielt das Reaktionsgemisch 75 ng *Hind*III+0 Primer, 75 ng *Mse*I+0 Primer, 0,2 mM dNTPs, 1/10 Vol. 10x Williams Puffer, 5 U Taq-Polymerase (Bioline). Die PCR-Bedingungen sind Tabelle 2.6 zu entnehmen.

Die selektive Amplifikation wurde in einem Gesamtvolumen von 10 μ l mit folgen-

Tabelle 2.6: Reaktionsbedingungen der AFLP Präamplifikation

Anzahl der Zyklen	Reaktionsschritt	Zeit	Temperatur
1	Anfangsdenaturierung	5 min	94 °C
20	Denaturierung	30 sek	94 °C
	Hybridisierung	30 sek	60 °C
	Synthese	1 min	72 °C
1	Endsynthese	10 min	72 °C

den Reagenzien durchgeführt: 2,5 Vol. präamplifizierte DNA (1:20 in 1x TE-Puffer verdünnt), 4 ng *Hind*III-Primer (700 IRD markiert), 4 ng *Hind*III-Primer (800 IRD markiert), 15 ng *Mse*I-Primer, 0,2 μ M dNTPs, 0,25 U Taq-Polymerase, 1/10 Vol. 10x Williams Puffer. Die *Hind*III-Primer sind mit einem Farbstoff markiert, der bei einer Wellenlänge von 700 oder 800 nm vom Laser des LiCor® Sequenzer Systems detektiert werden kann. Die Selektivität der Amplifikation wurde durch das Anhängen von drei (*Hind*III-Primer) bzw. vier (*Mse*I-Primer) zusätzlichen Basen an das 3'-Ende bestimmt. Zur Analyse der Nachkommenschaft wurden 89 Primerkombinationen und zur Analyse der Genotypensammlung wurden 25 Primerkombinationen verwendet. Die Sequenzen und Primerkombinationen sind im Anhang 5.3 aufgeführt. Die PCR-Bedingungen für die selektive Amplifikation sind in Tabelle 2.7 dargestellt.

Das Produkt der selektiven Amplifikation wurde mit Ladepuffer für PAA-Gele (Tab. 5.5) versetzt und mittels Polyacrylamid Gelelektrophorese (siehe Abschnitt 2.4.2) aufgetrennt.

Tabelle 2.7: PCR Bedingungen für AFLP-Endreaktion

Anzahl der Zyklen	Reaktionsschritt	Zeit	Temperatur
1	Anfangsdenaturierung	5 min	94 °C
1	Denaturierung	30 sek	94 °C
	Hybridisierung	30 sek	65 °C
	Synthese	1 min	72 °C
11	Denaturierung	30 sek	94 °C
	Hybridisierung	30 sek	65 °C -0,7 °C pro Zyklus
	Synthese	1 min	72 °C
24	Denaturierung	30 sek	94 °C
	Hybridisierung	30 sek	56 °C
	Synthese	1 min	72 °C
1	Endsynthese	10 min	72 °C

2.5 Bioinformatik

Die Internetadressen der frei verfügbaren Software, sowie die Vertreiber der kommerziell erwerblichen Programme sind im Anhang aufgeführt.

2.5.1 Screening der Sequenzen nach Mikrosatellitenmotiven

Um in den von GATC Biotech AG ermittelten Sequenzen der angereicherten DNA Fragmente Mikrosatellitenmotive zu finden, wurde die frei erhältliche Software 'Tandem Repeats Finder' von Benson (1999) mit folgenden Einstellungen verwendet: Alignment Parameter: 2, 7, 7; Minimum Alignment Score: 30; Maximum Period Size: 50.

2.5.2 Design von Primern

Zur Konstruktion der Primerpaare wurde das frei erhältliche Programm 'Primer 3' von Rozen and Skaletsky (2000) mit den vorgegebenen Einstellungen verwendet.

2.5.3 Erstellung der genetischen Karte

Die genetische Karte wurde mit der vorhandenen Nachkommenschaft der Kreuzung '21x19' erstellt.

Für die Berechnung der genetischen Karte wurde die Software JoinMap 4 (VanOoijen (2006)), mit folgenden Modifikationen verwendet. Die Seggregationsdaten der Marker wurden als 'cross population' (CP) behandelt. Es wurden für jeden Elter unabhängig eine separate Karte erstellt. In die Berechnung sind nur Single Dose Marker (SDM) eingeflossen, die auch das entsprechende Spaltungsmuster aufwiesen (Kapitel 1.5). Für die Gruppierung der Marker ('Grouping') wurde eine Rekombinationsfrequenz von 0,6 und ein 'Independence'-LOD von 1 gewählt. In die Berechnung der maternalen Karte sind alle SDM eingeflossen, die von der Mutter (21) oder von beiden Elternteilen vererbt werden. Analog wurden für die Berechnung der paternalen Karte SDM verwendet, die von dem Vater (19) oder von beiden Elternteilen vererbt wurden. Für die Einteilung der Kopplungsgruppen wurde als unteres Signifikanzniveau ein LOD Score von 3 angenommen. Die Reihenfolge der Marker auf den Kopplungsgruppen wurde anschließend mit Hilfe der Kartierungsfunktion von Kosambi (1944) und der Methode der Regressionskartierung von Stam (1993) ermittelt. Marker, die im 'Grouping' nicht sofort in eine Kopplungsgruppe eingeordnet werden konnten, wurden manuell den entsprechenden Kopplungsgruppen zugewiesen. Dabei wurde als Grenze ein SCL-Wert von 2.0 angenommen. Der 'Strongest Cross Link'

(SCL) Parameter ermöglicht die Zuordnung von Markern in eine bereits existierende Kopplungsgruppe, selbst bei schwachen Kopplungen (VanOoijen (2006)). Bei der Regressionskartierung werden die einzelnen Marker nacheinander hinsichtlich ihrer besten Position ('goodness-of-fit') auf der genetischen Karte platziert. In der 'ersten Runde' werden Marker von der Karte entfernt, wenn der 'goodness-of-fit' zu stark abnimmt oder wenn es aufgrund dieses Locus zu negativen Distanzen kommen würde. Die Berechnung wird in der 'zweiten Runde' wiederholt und aufgrund der höheren Anzahl an paarweisen Daten können zuvor von der Karte entfernte Marker auf der Karte platziert werden. In der 'dritten Runde' werden alle Marker, die in den ersten beiden Runden nicht kartiert werden konnten, unabhängig von ihrer Abnahme der 'goodness-of-fit' und eventuell negativen Distanzen, auf der Karte platziert. Dementsprechend sind die Karten der 'ersten Runde' vorzuziehen. Die Karten der 'dritten Runde' dienen dazu, eine Vorstellung über die ungefähre Lage weniger gut gekoppelter Marker zu erhalten. Dabei kann die Reihenfolge der Marker allerdings nicht als sicher angenommen werden (VanOoijen (2006)).

Folgende Einstellungen wurden bei der Regressionskartierung nach Kosambi (1944) in Join Map 4 (VanOoijen (2006)) verwendet: Kopplungen zwischen zwei Markern wurden mit einer Rekombinationsfrequenz kleiner als 0,45 und einem LOD größer als 1 angenommen. Des Weiteren wurde die 'Goodness-of-fit jump threshold' auf 5 gesetzt. Alle anderen Einstellungen wurden nicht verändert.

Die Abdeckung des Genoms mit den kartierten Markern kann nach der von Chakravarti et al. (1991) entwickelten und in Stein et al. (2007) beschriebenen Methode nach folgenden Formeln berechnet werden:

$$E(G) = \frac{n(n-1)d}{k} \quad (2.3)$$

$E(G)$ ist die geschätzte Genomgröße der zu kartierenden Population, wobei n die Anzahl der kartierten Marker, d die maximale Distanz zwischen zwei Markern und k die Anzahl der Markerpaare angibt.

$$E(C) = \frac{l \cdot 100}{E(G)} \quad (2.4)$$

$E(C)$ gibt die geschätzte Abdeckung des Genoms mit den vorhandenen Markern auf der genetischen Karte an und wird durch das Verhältnis zwischen geschätzter Genomgröße und Kartenlänge (l) bestimmt.

2.5.4 QTL-Kartierung

Für die Berechnung des quantitativen Einflusses chromosomaler Regionen auf die zu untersuchende Eigenschaft wurde das Programm MapQTL 6 (VanOoijen (2009)) genutzt. Dazu wurden die Mittelwerte der morphologischen Daten verwendet. Untersucht wurden der Anthocyangehalt der Blätter, der Anthocyangehalt in den Tepalen, der Anthocyangehalt in der Lippe, die Blütengröße sowie die Rispenlänge. Zunächst wurden die morphologischen Daten mittels Kolmogorov-Smirnov-Tests in WinStat auf Normalverteilung getestet. Für nicht normalverteilte Daten wurde in WinStat eine BoxCox Transformation [$Y=\ln(0)$] nach Yang et al. (2006) durchgeführt. Nach der Transformation wurden die Daten erneut auf Normalverteilung getestet.

Um die Signifikanzschwelle für QTL festzulegen, wurde ein Permutationstest nach Churchill and Doerge (1994) mit 1 000 Wiederholungen durchgeführt. Als Signifikanzschwelle wurde das 95%ige Konfidenzintervall jeweils als genomweite Schwelle für die maternale und paternale Karte, sowie als chromosomen- bzw. kopplungsgruppenweite Schwelle für die einzelnen Kopplungsgruppen festgelegt. Für das Intervallmapping wurden beide in der Software implizierten Algorithmen (Regressions- und 'Mixture Model') getestet. Beide Algorithmen brachten ähnliche Ergebnisse und im Folgenden werden nur die Ergebnisse des 'Mixture Models' betrachtet. Hierbei soll der Einfluss chromosomaler Regionen, zwischen zwei Markern untersucht werden. Um den Einfluss einzelner Marker auf das Merkmal zu untersuchen, wurde der in dem Programm zur Verfügung stehende Kruskal-Wallis-Test¹ gerechnet. Hierbei wurde die Signifikanz des Tests durch eine Semibonferroni-Adjustierung erhöht. Dazu wurde der kritische Wert durch die Anzahl der Chromosomen im haploiden Genom geteilt ($\alpha=0,05/19$). Eine genauere Angabe des Einflusses des Markers ist in MapQTL 6 durch die Angabe der Teststatistik als K^* gegeben. Dazu wurde der kritische Wert in R berechnet [$qchisq(1-0.05/19, 1)$]. Der Freiheitsgrad ist $df=1$, da alle Marker dominant ausgewertet wurden.

Der Einfluss chromosomaler Regionen auf die Ausprägung des zu untersuchenden Merkmals kann erst dann als signifikant erklärt werden, wenn in diesem Intervall die zuvor festgelegte Signifikanzschwelle aus dem Permutationstest überschritten wurde und die in dieser Region befindlichen Marker signifikante, bonferroni-adjustierte K^* -Werte für den Kruskal-Wallis-Test zeigen.

¹Der Kruskal-Wallis Test ist eine Form der Varianzanalyse für Rangdaten. Es handelt sich hierbei um eine Erweiterung des U-Tests, bei der keine Voraussetzungen an die Verteilung der Population gestellt werden.

2.5.5 Phylogenetische Analysen

Für die phylogenetischen Analysen wurden die AFLP und Mikrosatelliten Marker dominant per Hand ausgewertet. Dazu wurde eine 0/1 Matrix erstellt, wobei 0 für eine nicht vorhandene Bande und 1 für eine vorhandene Bande kodiert. Konnte nicht eindeutig bestimmt werden, ob die Bande vorhanden ist, oder ist die Gesamtreaktion eines Genotypen ausgefallen, so wurde dies mit 2 kodiert.

Die Berechnung des Phenogramms erfolgte mit der frei erhältlichen Software PHYLIP 3.68 (Felsenstein (2005)). Zunächst wurden die Distanzen nach der, im Programm implizierten Methode von Nei and Li (1979) berechnet. Anschließend wurde aus diesen Distanzen das Phenogramm nach der von Saitou and Nei (1987) erstmals beschriebenen Neighbour-joining Methode konstruiert. Um die Zweige des berechneten Neighbour-joining Baumes abzusichern, wurden die Bootstrap Werte über 100 Wiederholungen nach Felsenstein (1985) berechnet. Die Darstellung des Neighbour-joining Baumes erfolgte mit der Software TreeView 1.6.6 (Page (1996)).

2.5.6 Bestimmung/Ermittlung der Populationsstruktur

Die Populationsstruktur der Genotypensammlung wurde mit Hilfe der frei erhältlichen Software STRUCTURE 2.2 (Pritchard et al. (2000) und Falush et al. (2007)) auf Grundlage der AFLP Marker ermittelt.

Das Programm benutzt einen Bayesischen Kluster Ansatz mit der Annahme, dass die Marker Loci nicht gekoppelt sind und sich innerhalb einer Population im Hardy-Weinberg Gleichgewicht befinden (Pritchard et al. (2000)). Vereinfacht gesprochen soll eine Populationsstruktur ermittelt werden, welche von den genotypischen Daten (Marker) als Zufallsstichprobe, der(unbekannten) Herkunft der Individuen und der(unbekannten) Allelfrequenz abhängt.

Um die Individuen auf Basis von Multilocus Markern in Subpopulationen zu gruppieren benutzt das Programm STRUCTURE 2.2 einen 'Markov Ketten Monte Carlo' (MCMC) Ansatz (Pritchard et al. (2000)). Hierbei handelt es sich um eine Klasse von Algorithmen, welche eine zufällige Stichprobe aus einer Wahrscheinlichkeit simuliert. Dabei ist die Methode abhängig von der Annäherung der Markov Kette an die stationäre Verteilung (Holsinger et al. (2002)). Bei der Markov Kette werden bestimmte Situationen/Zustände simuliert, die nur vom gegenwärtigen Zustand abhängen und nicht von vergangenen oder zukünftigen. Dabei spielt die Anzahl der 'Kettenglieder' eine große Rolle. So muss der Berechnung ein ausreichender Vorlauf (BurnIn) erlaubt werden, damit die Kette die stationäre Phase erreichen kann. In dieser Phase folgen die auftretenden Zustände tendenziell einer Wahrscheinlichkeitsverteilung. Es werden nun in bestimmten Intervallen Proben gezogen, aus denen die wahrscheinlichste Ver-

teilung berechnet wird. Das zweite MC steht in dieser Methode für 'Monte Carlo'. Dieser Zufallsalgorithmus sorgt dafür, dass ausgehend vom vorherigen Kettenglied und bestimmten Regeln das nächste Kettenglied akzeptiert wird oder nicht. Durch diese Methode wird die Wahrscheinlichkeit einer Falschaussage, mit zunehmender Anzahl an Generationen, reduziert.

Für die Berechnung der Populationsstruktur wurden folgende Parameter gewählt: Aufgrund der Züchtungsgeschichte von *Phalaenopsis* Hybriden ist davon auszugehen, dass es zu einer Durchmischung des Genmaterials kam und Genotypen folglich Genmaterial unterschiedlicher Subpopulationen enthalten können. Demzufolge wurde das in STRUCTURE 2.2 implizierte 'Admixture Model' gewählt. Um die AFLP Marker dominant auszuwerten, wurde entsprechend dem Handbuch der Software STRUCTURE 2.2 ein rezessives Allel eingefügt. Obwohl vermutet wird, dass die untersuchten Genotypen unterschiedliche Ploidiestufen aufweisen, wurde als Ploidielevel 'diploid' angegeben, da dies die einzig mögliche Einstellung bei dominanten Markern in dem Programm ist. Nachdem festgestellt wurde, dass die Schwankungen der Ergebnisse, bei einem BurnIn von 10 000 und einer MCMC Wiederholung von 90 000 Generation, relativ hoch waren, wurde die Anzahl der BurnIns auf 100 000 und die der Generationen auf 900 000 erhöht. Die Simulation wurde für $K=2$ bis $K=15$ durchgeführt, wobei K die Anzahl der vermuteten Subpopulationen darstellt. Für jedes K wurde die Simulation unabhängig voneinander viermal durchgeführt. Nach jedem Lauf erhält man einen Wert $\ln P(D)$, der den 'geschätzten \ln der Wahrscheinlichkeit der Daten' angibt. Mit diesem Wert kann nun die wahrscheinliche Anzahl der Subpopulationen berechnet werden. In der vorliegenden Arbeit wurden dazu drei verschiedene Methoden getestet.

2.5.6.1 Berechnung der Anzahl der Subpopulationen nach Pritchard et al. (2000)

Zunächst wurde die Anzahl der Subpopulation berechnet, wie in Pritchard et al. (2000) und im Handbuch der Software STRUCTURE 2.2 beschrieben. Die Berechnung basiert auf einem Vergleich der für die jeweils angenommene Anzahl an Subpopulation von STRUCTURE 2.2 berechneten $\ln P(D)$ Werte. Ein Beispiel für die Berechnung ist dem Handbuch der Software STRUCTURE 2.2 auf Seite 13 zu entnehmen. Die Berechnungen wurden in MS Excel durchgeführt und die Ergebnisse sind in Tabelle 3.8 zu finden. Die Anzahl der Subpopulationen ist dort zu wählen, wo die geschätzte posterior Wahrscheinlichkeit einen maximalen Wert annimmt (0-minimaler Wert, 1-maximaler Wert).

2.5.6.2 Berechnung der Anzahl der Subpopulationen nach Cockram et al. (2008)

Die zweite Methode zur Bestimmung der Anzahl an Subpopulationen ist in Cockram et al. (2008) beschrieben. Die Protokolle dieser Methode wurden von J. Cockram (persönliche Mitteilung) zur Verfügung gestellt.

Äquivalent zur ersten Methode wird auch hier die Anzahl an Subpopulationen aufgrund des maximalen $\ln P(D)$ -Wertes geschätzt. Hier wird dieser Wert jedoch nicht berechnet, sondern auf Grundlage einer graphischen Darstellung geschätzt. Die von STRUCTURE 2.2 ausgegebenen Werte, werden in Abhängigkeit der Anzahl der Subpopulationen aufgetragen (siehe Abbildung 3.20).

Ebenfalls gibt STRUCTURE 2.2 so genannte 'Q-Matrizen' aus. Hier werden für jede Berechnung die Wahrscheinlichkeiten der Zugehörigkeit jedes Genotyps zu einer bestimmten Subpopulation dargestellt. Um die Q-Matrizen für jede Berechnung miteinander zu vergleichen, werden die Korrelationen zwischen den einzelnen Matrizen berechnet und aus den maximalen Korrelationswerten der Mittelwert gebildet. Diese Werte werden anschließend in Abhängigkeit der Anzahl der Subpopulationen aufgetragen (siehe Abbildung 3.21). Die Berechnungen wurden in MS Excel durchgeführt.

Die Anzahl der Subpopulationen ist bei dem K zu wählen, wo sich der $\ln P(D)$ -Wert in der Plateauphase befindet und der Vergleich der Q-Matrizen eine maximale Korrelation zeigt.

2.5.6.3 Berechnung der Anzahl der Subpopulationen nach Evanno et al. (2005)

Der dritte Ansatz zur Identifizierung der Anzahl an Subpopulationen erfolgte nach Evanno et al. (2005). Wie in der Veröffentlichung wird auch in dieser Arbeit ' $\ln P(D)$ ' bei diesem Ansatz als $L(K)$ bezeichnet.

Es werden pro Durchlauf vier Graphen erzeugt in denen nacheinander folgende Werte in Abhängigkeit der Anzahl der Subpopulationen (K) aufgetragen werden. In der ersten Graphik der $L(K)$ -Wert. In der zweiten Graphik die Rate der Änderung, berechnet nach $L'(K) = L(K) - L(K-1)$. In der dritten Graphik die absoluten Werte der zweiten Rate der Änderung, berechnet nach $|L''(K)| = |L'(K+1) - L'(K)|$. In der vierten Graphik ΔK , berechnet mit der Formel $\Delta K = m|L''(K)| / s[L(K)]$.

Die Anzahl der Subpopulationen ist bei dem K zu wählen, wo das maximale ΔK erreicht wird.

2.5.7 Ermittlung des Kopplungsungleichgewichts

Zur Berechnung des Kopplungsungleichgewichts wurde das Programm TASSEL 2.1 verwendet. Marker, die auf den genetischen Karten lokalisiert werden konnten, wurden in Abhängigkeit zu ihrem gegenseitigen Abstand aufgetragen, um den Abfall des Kopplungsungleichgewichts im *Phalaenopsis* Genom abschätzen zu können. Um das Kopplungsungleichgewicht zwischen den für die Assoziationsstudie verwendeten Markern abzuschätzen wurde das Kopplungsungleichgewicht zwischen jedem Markerpaar in TASSEL 2.1 berechnet und in dem Programm graphisch dargestellt.

2.5.8 Berechnung der Marker-Eigenschafts-Assoziationen

Genomweite Assoziationsstudien unterliegen folgendem allgemeinen linearem Model (GLM):

$$Y = G + Q + \varepsilon.$$

Dabei sind Y die phänotypischen Merkmale, G die genomweiten Polymorphismen (=Markerdaten), Q die Populationsstruktur und ε der Restfehler.

Bei den hier verwendeten Methoden handelt es sich um 'Einzel Marker Eigenschafts Assoziationen', wobei jeder Marker einzeln gegen die Nullhypothese getestet wird. Die Hypothesen für die folgenden Tests sind folgendermaßen formuliert:

H_0 : Marker hat keinen Einfluss auf das untersuchte Merkmal (no association)

H_A : Marker hat einen signifikanten Einfluss auf das untersuchte Merkmal

2.5.8.1 Assoziationsstudie in R

Die Assoziationsstudie wurde in R 2.8.1 durchgeführt. Dazu wurde eine Routine verwendet, welche von J. White (2009) geschrieben und freundlicherweise zur Verfügung gestellt wurde. Der Code ist dem Anhang (5.6) zu entnehmen. Die in STRUC-TURE 2.2 berechnete Populationsstruktur wird hier als Kovariate verwendet. Es werden vier verschiedene Berechnungen ausgegeben:

- | | |
|----------------------|---|
| Raw.Association.P | - gibt die Assoziationen zwischen Marker und Eigenschaft wieder, ohne die Populationsstruktur zu beachten |
| S.Association.P | - gibt die Assoziationen zwischen Marker und Eigenschaft wieder, bezieht hier die Populationsstruktur als Kovariate in die Berechnungen mit ein |
| Raw.Association.GC.P | - gibt die Assoziationen zwischen Marker und Eigenschaft |

wieder, ohne die Populationsstruktur zu beachten, bezieht aber eine Korrektur (Genomic Control) nach Devlin et al. (2000) mit ein

S.Association.GC.P - gibt die Assoziation zwischen Marker und Eigenschaft wieder, bezieht die Populationsstruktur als Kovariate sowie die Korrektur (Genomic Control) nach Devlin et al. (2000) in die Berechnung mit ein

Die Berechnungen erfolgen nach einem in R implizierten GLM Modell, welches einen einfachen 'Likelihood-Ratio-Test' benutzt, um die Nullhypothese zu testen. Die berechneten p-Werte werden -Log10 transformiert und in einem x/y-Diagramm dargestellt. Als Signifikanzschwelle wird $p=0,05$ angenommen. Ein höheres Signifikanzniveau wird durch eine Bonferroni-Adjustierung erreicht, indem der Schwellenwert durch die Anzahl der untersuchten Marker geteilt wird. Für die vergleichende Darstellung mit den berechneten p-Werten im x/y-Diagramm wird der adjustierte p-Wert ebenfalls -Log10 transformiert und als waagerechte Linie im Diagramm dargestellt. Berechnete p-Werte, die über dieser Linie liegen, zeigen Marker, die eine signifikante Assoziation zu der untersuchten Eigenschaft aufweisen.

2.5.8.2 Assoziationsstudien in TASSEL 2.1

Das Kopplungsungleichgewicht zwischen den einzelnen Markern sowie die Assoziationen zwischen Marker und phänotypischer Eigenschaft wurden mit TASSEL 2.1 (Bradbury et al. (2006)) berechnet. Für die Berechnung des Kopplungsungleichgewichts wurden 1 000 Wiederholungen gewählt. Die Berechnung der Assoziationen erfolgte mit dem in TASSEL 2.1 integrierten GLM Modell. Hierbei wird für jede Marker-Eigenschafts Kombination die Lösung nach der allgemeinen statistischen Methode der kleinsten Quadrate ('least squares method') gefunden (Bradbury et al. (2006)). Dabei wurde die Q-Matrix aus STRUCTURE 2.2 als Kovariate ins Modell mit einbezogen.

2.5.9 Graphische Darstellung

Die Darstellung der genetischen Karten und QTL erfolgte durch JoinMap 4 oder durch die Software MapChart 2.2 (Voorrips (2002))

Die Darstellung der Graphiken erfolgte, soweit nicht anders angegeben, mit dem Graphikpaket 'graphics' von R 2.8.1 (R Team (2008)).

3 Ergebnisse

3.1 Kartierung gartenbaulich wichtiger Merkmale

3.1.1 Phänotypische Merkmale

Die Genotypen 19 (Blütenfarbe: weiß) und 21 (Blütenfarbe: violett) wurden reziprok gekreuzt. Dabei zeigte sich eine maternale Vererbung der Blütenfarbe. Die Nachkommen der Kreuzung 19x21 zeigten ausschließlich weiße Blüten mit höchstens einem Hauch rosa in der Färbung. Die Nachkommen der Kreuzung '21x19' zeigten hingegen violette Blüten in unterschiedlichen Schattierungen (siehe Abbildung 3.1 und 3.2).

Für die vorliegende Arbeit wurde die Kreuzung '21x19' untersucht. In Abbildung 3.1 bis 3.3 sind die Unterschiede in den phänotypischen Ausprägungen der Nachkommenschaft '21x19' dargestellt. Ein höherer Anthocyangehalt spiegelte sich in einem höheren OD-Wert wieder. Der Anthocyangehalt variierte von 0,0113 bis 0,187 OD in den Tepalen, von 0,0118 bis 0,2408 OD in den Blättern und von 0,0107 bis 0,615 OD in der Lippe. Der Blütendurchmesser der Nachkommenschaft variierte von 4 bis 6,3 cm und die Rispenlänge reicht von 17 bis 76 cm.

Tabelle 3.1: Kolmogorov-Smirnov-Test der phänotypischen Daten

Test auf Normalverteilung (Kolmogorov-Smirnov-Test für kontinuierliche Variable)			
	N	D	p-Wert
Anthocyangehalt in der Blüte	55	0,090504448	0,758488195
Anthocyangehalt im Blatt	87	0,128239977	0,114342683
Anthocyangehalt in der Lippe	55	0,192631007	0,033756066
Blütengröße	53	0,145208196	0,21371057
Rispenlänge	54	0,103485582	0,609587713

In Abbildung 3.4 sind die Histogramme der Boniturdaten und ihre Verteilung dargestellt. Da aus diesen Abbildungen nicht eindeutig ersichtlich wird, ob die Daten normalverteilt sind, wurde mit Hilfe des Kolmogorov-Smirnov-Tests auf Normalverteilung getestet (siehe Tabelle 3.1). Alle Daten, bis auf ‚Anthocyangehalt in der Lippe‘ können als normalverteilt angenommen werden. Nicht normalverteilte Daten konnten



Abbildung 3.1: Variation der Blütenmorphologie in der Nachkommenschaft '21x19'



Abbildung 3.2: Unterschiede in der Ausprägung des Anthocyangehalts der Blüten in der Nachkommenschaft '21x19'



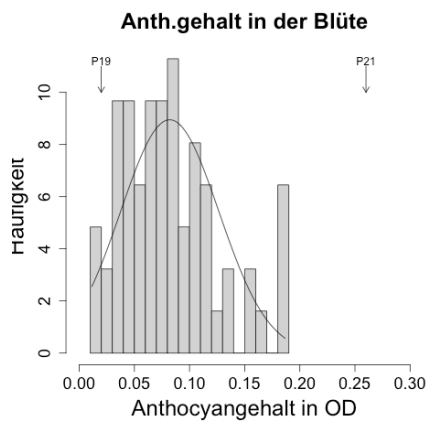
(a) Blütengröße



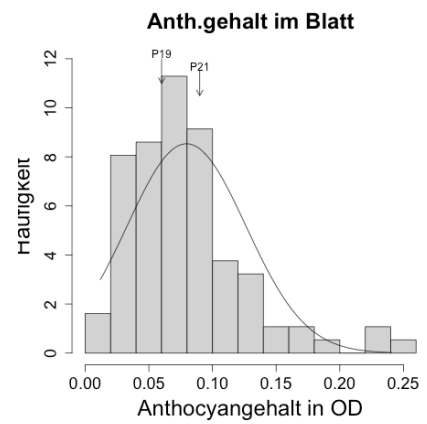
(b) Rispenlänge

Abbildung 3.3: Unterschiede in der Blütengröße (a) und der Rispenlänge (b) in der Nachkommenschaft der Kreuzung '21x19'

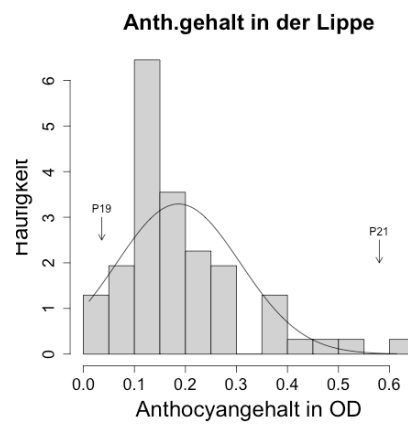
mit Hilfe der BoxCox-Transformation in normalverteilte Daten transformiert werden. Nach der Transformation ergab der Kolmogorov-Smirnov-Test einen p-Wert von $p=0,1033$ für die Daten ,Anthocyangehalt in der Lippe.



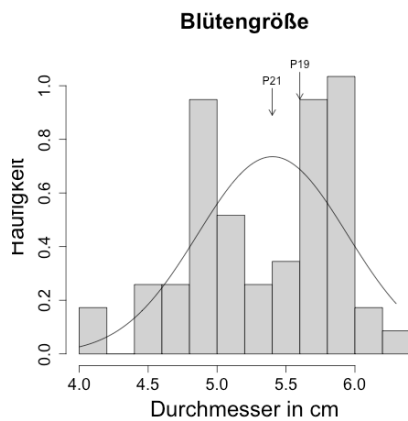
(a)



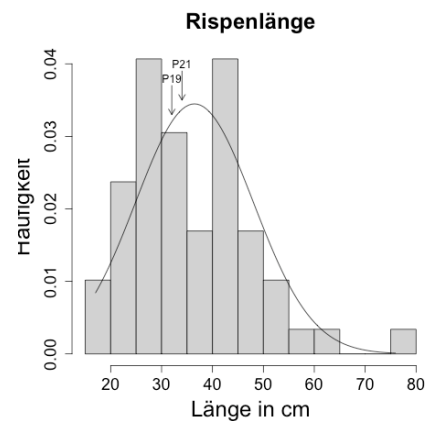
(b)



(c)



(d)



(e)

Abbildung 3.4: Histogramme der untersuchten phänotypischen Merkmale der Nachkommenschaft '21x19': (a) Anthocyangehalt in den Tepalen, (b) Anthocyangehalt im Blatt, (c) Anthocyangehalt in der Lippe, (d) Blütengröße, (e) Rispenlänge; die phänotypischen Ausprägungen der Eltern sind mit 'P19' und 'P21' gekennzeichnet

3.1.1.1 Zusammenhang zwischen phänotypischen Merkmalen

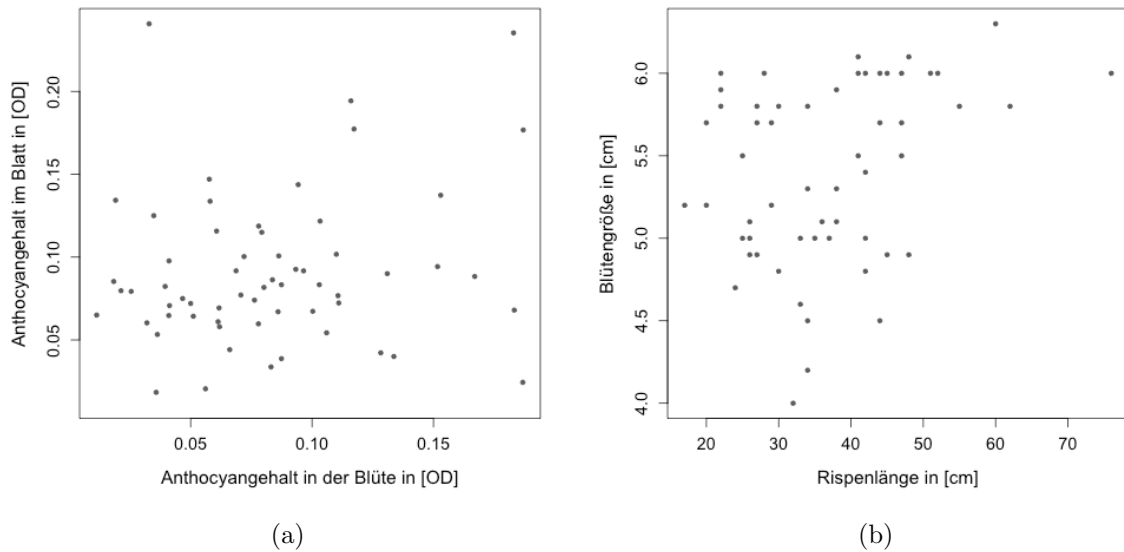


Abbildung 3.5: Zusammenhang zwischen der Ausprägung unterschiedlicher Merkmale der Individuen der Nachkommenschaft '21x19': (a) Graphische Darstellung des Anthocyangehaltes im Blatt (y-Achse) und des Anthocyangehalt in der Blüte (x-Achse); (b) Graphische Darstellung der Blütengröße (y-Achse) und der Rispenlänge (x-Achse)

Um Zusammenhänge zwischen dem Anthocyangehalt im Blatt und in der Blüte sowie Zusammenhänge zwischen der Blütengröße und Rispenlänge zu detektieren, wurde die Korrelation zwischen den entsprechenden Merkmalen berechnet. In Abbildung 3.5 (a) ist der Anthocyangehalt des Blattes in Abhängigkeit zum Anthocyangehalt der Blüte aufgetragen. Der Korrelationskoeffizient betrug 0,1785. In Abbildung 3.5 (b) ist die Rispenlänge in Abhängigkeit der Blütengröße aufgetragen. Der Korrelationskoeffizient betrug 0.3182.

3.1.2 Mikrosatellitenmarker

3.1.2.1 Entwicklung von Mikrosatellitenmarkern

Die Entwicklung von Mikrosatellitenmarkern wurde in Abschnitt 2.4.3, 2.4.4 und 2.5.1 beschrieben. Um möglichst viele Fragmente für die Entwicklung von Mikrosatellitenmarkern zur Verfügung zu haben, wurden die in diesen Abschnitten beschriebenen Schritte zweimal durchgeführt. Die Ergebnisse der Mikrosatellitenanreicherung und -testung sind in Tabelle 3.2 aufgeführt.

Die Ergebnisse in dieser Tabelle sind in zwei Abschnitte unterteilt, in denen jeweils die Ergebnisse des jeweiligen Experiments aufgeführt sind (1. und 2. Durchgang). Die DNA-Fragmente wurden nach den Motiven AG und GT angereichert und jeweils nach der zweiten und dritten Anreicherungsrunde kloniert. In Klammern ist die Gesamtzahl der Klone aufgeführt, die als positive Klone gepickt wurden. Der absolute und prozentuale Anteil der Klone mit positivem Hybridisierungssignal ist ebenfalls in der entsprechenden Zelle aufgeführt. Die Anzahl positiver Klone zeigt die Summe der Klone mit positivem Hybridisierungssignal für das jeweilige Anreicherungsmotiv. Insgesamt zeigten von 3267 gepickten Klonen 441 Klone ein positives Hybridisierungssignal. Von diesen wurden insgesamt 160 Klone sequenziert. In den Sequenzen (nur zwei Sequenzen waren nahezu identisch) konnten insgesamt 137 Mikrosatellitenmotive detektiert werden. Aus den Mikrosatellitenmotiv-enthaltenden Sequenzen konnten 127 Primerpaare konstruiert werden, welche anschließend direkt im 6%igen Polyacrylamidgel ('PAA') nach der in Abschnitt 2.4.4 beschriebenen Methode getestet wurden. Ein Primerpaar galt als funktional, wenn es klare eindeutige Amplifikate bildete. Als Testgenotypen wurden die Genotypen 0744, 1307, 19, 21, *Ph. violacea*, *Ph. hieroglyphica*, 0739, 0771, 0774, 0781, 0823, 0836, 5702 und 5710 verwendet. Monomorphe Primerpaare ('mono') zeigten keine Unterschiede im Bandenmuster dieser vierzehn unterschiedlichen Genotypen. Polymorphe Primerpaare ('poly') zeigten Unterschiede im Bandenmuster in mindestens einem der vierzehn Genotypen.

Da die Fragmente für die Motive (AG)₁₀ und (GT)₁₀ angereichert wurden, wurde vorrangig nach diesen Motiven gesucht. Bei der Suche nach diesen Motiven wurden jedoch auch andere Motive gefunden, für die ebenfalls Primerpaare entwickelt wurden. In Tabelle 5.8 sind alle Primer und das entsprechende Mikrosatellitenmotiv, welches sie amplifizieren aufgelistet, die in Tabelle 3.2 als polymorph deklariert wurden und für die weitere Analysen verwendet wurden. Insgesamt wurden 17 funktionale polymorphe Mikrosatellitenmarker für *Phalaenopsis* Hybriden entwickelt.

Tabelle 3.2: Übersicht der Effektivität der Entwicklung von Mikrosatellitenmarker; von der Erstellung einer angereicherten DNA-Bank bis zur Konstruktion der Primerpaare

1. Durchgang									
Motiv	Anreicherungs-effizienz ¹				Primereffizienz				
	2. Anreicherungs- runde	3. Anreicherungs- runde	Anzahl positive Klone	Anzahl seq. Klone	Anzahl SSR - Motive	Duplikate	Anzahl konstr. Primerpaare	Primerpaare funktional auf PAA ²	
AG	64 (408) =15,7%	116 (687) =16,9%	180	66	66	1	73	(21 mono ³) 10 poly ⁴	
GT	20 (286) =7%	48 (482) =9,9%	68	26	22	0	10	1 poly	
Gesamt	84	164	248	92	88	1	83	11	

2. Durchgang									
Motiv	Anreicherungs-effizienz ¹				Primereffizienz				
	2. Anreicherungs- runde	3. Anreicherungs- runde	Anzahl positive Klone	Anzahl seq. Klone	Anzahl SSR - Motive	Duplikate	Anzahl konstr. Primerpaare	Primerpaare funktional auf PAA ²	
AG	75 (496) =15,1%	31 (274) =11,3%	106	27	24	0	32	(12 mono) 4 poly	
GT	14 (148) =9,5%	73 (486) =15%	87	41	25	0	12	(1 mono) 2 poly	
Gesamt	89	104	193	68	49	0	44	6	

¹ Anreicherungs-effizienz berechnet sich als das prozentuale Verhältnis zwischen der Anzahl der gepickten Klone (in Klammern dargestellt) und der Anzahl der Kolonien die ein positives Hybridisierungssignal zeigten

² PAA - Polyacrylamidgel

³ mono - Primerpaare erzeugen monomorphe Amplifikate

⁴ poly - Primerpaare erzeugen polymorphe Amplifikate

3.1.2.2 Anwendung der Mikrosatellitenmarker

Neben den aus Genotyp 0744 entwickelten Mikrosatelliten wurden fünf im Internet veröffentlichte Mikrosatelliten verwendet. Die Primer sowie deren Quelle sind in Tabelle 5.10 (siehe Anhang) aufgeführt.

Insgesamt wurden 22 Mikrosatellitenprimerpaare für die in dieser Arbeit beschriebenen Analysen verwendet. Die Mikrosatelliten zeigten jedoch ein sehr komplexes Bandenmuster, welches in der Nachkommenschaft weder einem diploiden noch einem polyploiden Seggregationsmuster zugeordnet werden konnte. Des Weiteren konnte nicht eindeutig festgestellt werden, ob ein oder mehrere Loci amplifiziert wurden. Deshalb mussten die Mikrosatellitenmarker analog der AFLP-Marker dominant ausgewertet werden.

Bei der Anwendung der Mikrosatellitenmarker auf die Genotypensammlung erzeugten von 22 Mikrosatellitenprimerpaaren sechs ein auswertbares Bandenmuster. In Tabelle 3.4 sind die Mikrosatellitenmarker mit ihren PIC Werten aufgeführt. Dabei ist der Informationsgehalt eines Markers umso höher zu bewerten, je dichter der PIC Wert an 1 liegt.

Tabelle 3.4: Mikrosatellitenmarker und ihre PIC-Werte; ermittelt in einer Population bestehend aus 144 verschiedenen *Phalaenopsis* Genotypen

Marker	Anzahl Allele	Heterozygotie	PIC
MPH05	25	0,86	0,85
MPH27	7	0,78	0,74
MPH39	2	0,46	0,35
MPH51	8	0,83	0,80
MPH58	3	0,58	0,49
L3	18	0,86	0,84

3.1.3 Erstellung der genetischen Karte

Für die Erstellung einer genetischen Karte wurden Mikrosatelliten- und AFLP Marker verwendet.

Von 22 Mikrosatellitenprimerpaaren erzeugten zwölf ein polymorphes Bandenmuster, das eindeutig ausgewertet werden konnte:

MPH05(5), MPH27(4), MPH51(2), MPH62(2), MPH88(2), MPH120(5),
MPH123(4), L2(7), L3(6), L4(6), L5(2), S2(3)

(In Klammern ist jeweils die Anzahl der Allele aufgeführt.)

Zur Erzeugung von AFLP Markern wurden 89 Primerkombinationen auf die Nachkommenschaft der Kreuzung '21x19' mit 103 Individuen angewendet. Insgesamt wurden pro Primerkombination im Durchschnitt 77 Banden pro Genotyp amplifiziert, von denen durchschnittlich 8 Banden Polymorphismen zwischen den Genotypen zeigten und auswertbar waren. Diese Banden wurden, nach Ermittlung des Spaltungsverhältnisses, als Marker für die Erstellung der genetischen Karte verwendet.

Tabelle 3.5: Übersicht der Marker, die für die Erstellung der genetischen Karte verwendet wurden (in Klammern hinter der Markerbezeichnung ist die Gesamtzahl der detektierten Marker zu sehen)

Marker	AFLP (707)			SSR (dominant, 47)		
	paternal	maternal	biparental	paternal	maternal	biparental
detektiert	253	323	131	23	20	5
SDM ¹	169	199	70	18	12	0
kartiert	158	145	8	15	4	0

¹ SDM - Single Dose Marker (nähere Erläuterungen siehe Kapitel 1.5)

Für die Erstellung der genetischen Karte konnten insgesamt 707 AFLP-Marker und 47 dominante Mikrosatelliten-Marker genutzt werden. Davon wurden 465 AFLP- und 41 SSR-Marker als Single Dose Marker (SDM), mit einem Spaltungsverhältnis von 1:1 für maternal oder paternal vererbte und von 1:3 für biparental vererbte Marker, identifiziert. Insgesamt konnten 311 AFLP- und 19 Mikrosatelliten-Marker auf den genetischen Karten lokalisiert werden. In Tabelle 3.5 ist eine Zusammenfassung der detektierten und kartierten Marker aufgelistet.

3.1.3.1 Paternale Karte

In Abb 3.6 ist die paternale Karte dargestellt. Es konnten 19 Kopplungsgruppen mit insgesamt 181 Markern identifiziert werden. Dies entspricht der Anzahl der Chromosomen im haploiden Genom von *Phalaenopsis*. Die Gesamtlänge der paternalen Karte beträgt 1538,3 cM mit einer Durchschnittslänge von 81 cM je Kopplungsgruppe. Es befinden sich durchschnittlich 9,5 Marker mit einem durchschnittlichen Abstand von 8,5 cM auf jeder Kopplungsgruppe. Die größte Lücke befindet sich auf Kopplungsgruppe 9 zwischen den Markern H800ATT_MACTG_1 und H700CGA_MACAG_9 mit 41,7 cM. Die nach Formel 2.3 und 2.4 errechnete Abdeckung des Genoms beträgt 30,84%.

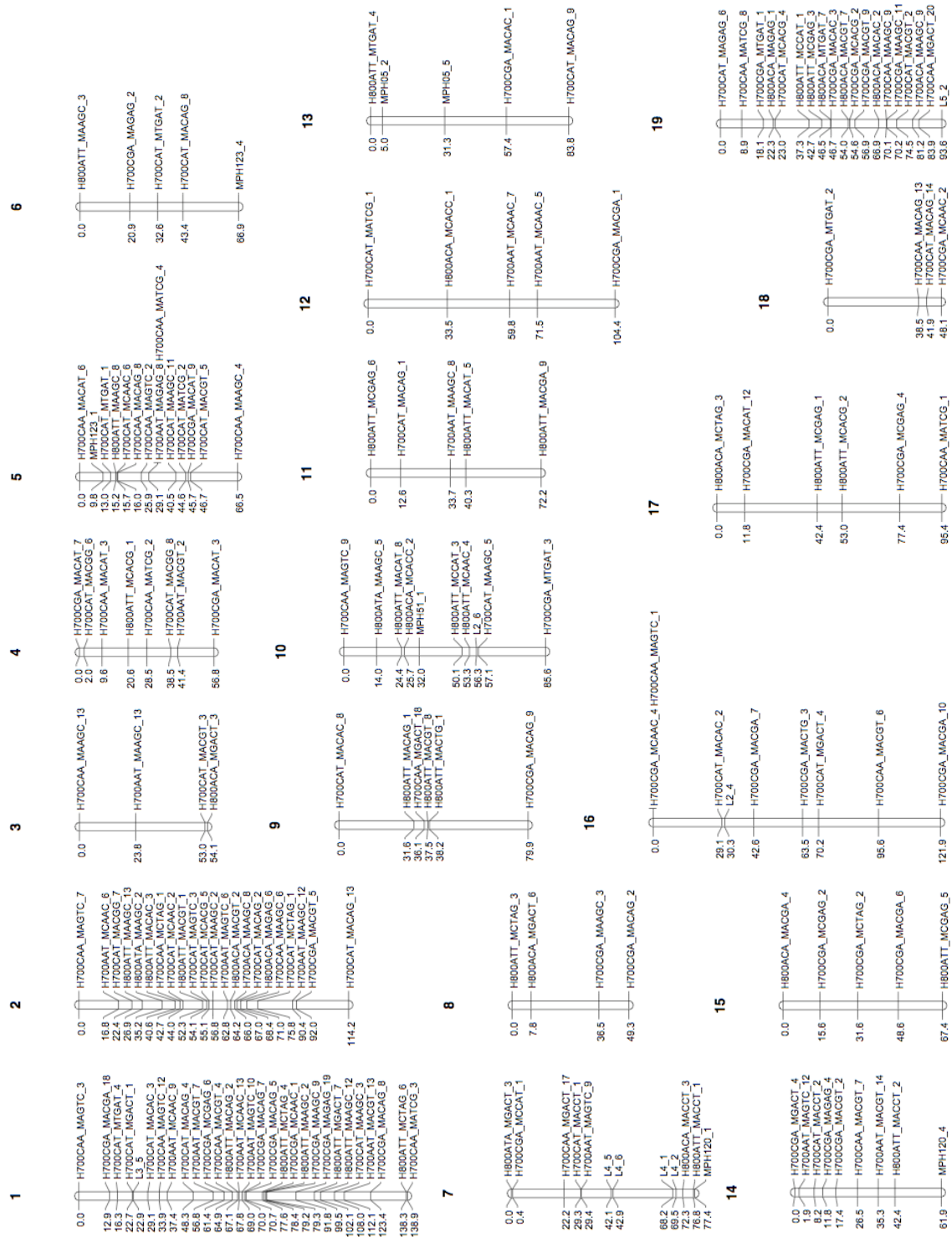


Abbildung 3.6: Übersicht der Kopplungsgruppen der paternalen Karte

3.1.3.2 Maternale Karte

In Abb 3.7 ist die maternale Karte dargestellt. Wie in der paternalen Karte konnten auch hier 19 Kopplungsgruppen mit insgesamt 157 Markern, entsprechend der Chromosomenzahl des haploiden *Phalaenopsis*-Genoms identifiziert werden. Die Gesamtlänge der maternalen Karte beträgt 1179,8 cM mit einer durchschnittlichen Länge von 62,1 cM je Kopplungsgruppe. Es befinden sich durchschnittlich 8,3 Marker mit einem durchschnittlichen Abstand von 7,5 cM auf jeder Kopplungsgruppe. Der größte Abstand befindet sich auf Kopplungsgruppe 17 zwischen den Markern H800ATT_MGACT_4 und H800ATT_MAAGC_5 mit 34,9 cM. Die nach Formel 2.3 und 2.4 errechnete Abdeckung des Genoms beträgt 29,22%.

Verknüpfung der Kopplungsgruppen

Eine Verknüpfung der Kopplungsgruppen der beiden Karten ist aufgrund einer zu geringen Anzahl an Ankermarkern nicht möglich. Biparentale Marker, die als Ankermarker dienen sind in Tabelle 3.6 aufgeführt.

Tabelle 3.6: Ankermarker

Marker	paternale Karte	maternale Karte
H800ATT_MACAC_3	LG2	LG4
H700AAT_MCAAC_6	LG2	LG4
H700CAA_MACAT_6	LG5	LG5
H700CGA_MAGAG_2	LG6	LG14
H700AAT_MAGTC_12	LG14	LG9
H700CAA_MACGT_7	LG14	LG5
H800ATT_MACCT_2	LG14	LG5
H800ACA_MACGA_4	LG15	LG13

LG - Kopplungsgruppe (linkage group)



Abbildung 3.7: Übersicht der Kopplungsgruppen der maternalen Karte

3.1.4 QTL-Analyse in der Nachkommenschaft

Die Nachkommenschaft der Kreuzung '21x19' wurde auf folgende Eigenschaften untersucht: Anthocyangehalt in der Blüte (Tepalen), Anthocyangehalt im Blatt, Anthocyangehalt in der Lippe, Blütengröße und Rispenlänge.

Die Mittelwerte der quantitativen Analysen, sowie die BoxCox transformierten Mittelwerte für den Anthocyangehalt in der Lippe (logLip) flossen in die QTL Berechnungen ein.

Die Signifikanzschwelle wurde, entsprechend einer semibonferroni Adjustierung (siehe Abschnitt 2.5.4) und einem Freiheitsgrad von $df=1$ für dominante Marker, in R 2.8.1 berechnet: $qchisq(1-0.05/19,1)=9,04678$. In Tabelle 3.7 sind alle Marker aufgeführt, die die Signifikanzschwelle von $K^*=9,04678$ überschreiten und somit einen signifikanten Einfluss auf das entsprechende Merkmal haben.

Die kopplungsgruppenweiten und genomweiten Signifikanzschwellen wurde mittels Permutationstest in MapQTL 6 berechnet und sind in Tabelle 5.14 aufgeführt. In den folgenden Abbildungen ist die kopplungsgruppenweite Signifikanzschwelle als gestrichelte Linie und die genomweite Signifikanzschwelle als durchgezogene Linie dargestellt.

Abbildung 3.8 zeigt die Ergebnisse der QTL-Analyse mit MapQTL 6 für das Merkmal Blütengröße. Es wurde eine Kopplungsgruppe gefunden, die einen Einfluss auf dieses Merkmal in der vorliegenden Nachkommenschaft zeigt. Zum einen zeigen die Marker H700CGA_MACGT_11 und H700CGA_MCAAC_5 einen signifikanten Einfluss im Kruskal-Wallis-Test. Und zusätzlich konnte im Intervallmapping die Kopplungsgruppenweite Signifikanzschwelle in dieser Region überschritten werden. Da auch die umliegenden Marker einen, wenn auch nur geringen, nicht signifikanten Einfluss zeigen, scheint diese Region einen starken Einfluss auf die Blütengröße zu haben.

In Abbildung 3.9 sind drei Kopplungsgruppen zu sehen, die einen signifikanten Einfluss auf das Merkmal Anthocyangehalt in der Lippe haben. In Kopplungsgruppe 2 und 6 sind keine Kruskal-Wallis signifikanten Marker zu finden. Jedoch zeigte das Intervallmapping, dass Regionen dieser Kopplungsgruppen einen Einfluss haben können, da die kopplungsgruppenweite Signifikanzschwelle überschritten wurde. Der Marker H700CAA_MATCG_8 auf Kopplungsgruppe 19 der paternalen Karte hat laut Kruskal-Wallis-Test einen signifikanten Einfluss auf den Anthocyangehalt in der Lippe. Auch die um diesen Marker liegende Region zeigte im Intervallmapping einen signifikanten Einfluss.

Abbildung 3.10 zeigt die Kopplungsgruppe 6 der paternalen Karte. Der Marker H800ATT_MAAGC_2 zeigt im Kruskal-Wallis-Test einen signifikanten Einfluss auf das Merkmal Anthocyangehalt in der Blüte. Auch im Intervallmapping konnte in der

Tabelle 3.7: **QTL in der Nachkommenschaft:** Marker, die nach dem Kruskal-Wallis-Test einen signifikanten Einfluss auf das entsprechende Merkmal in der Nachkommenschaft haben. Neben dem untersuchten Merkmal ist die Kopplungsgruppe (KG) und Position des Markers sowie dessen Teststatistik (K^*) aufgeführt. (U - Marker konnte nicht auf der Karte positioniert werden)

maternale Karte				
Merkmal	KG	Position	Marker	K^*
Anth.gehalt in der Blüte	keine signifikanten Marker			
Anth.gehalt im Blatt	U	11	L4.4	9,765
	19	39,093	H700CGA_MACGA_19	9,185
Anth.gehalt in der Lippe	keine signifikanten Marker			
Blütengröße	1	26,654	H700CGA_MACGT_11	10,011
	1	29,854	H700CGA_MCAAC_5	9,14
Rispenlänge	18	52,964	H700CAT_MACAT_3	10,672
	13	112,747	H700CAT_MCCAT_8	8,549
loglip	keine signifikanten Marker			

paternale Karte				
Merkmal	KG	Position	Marker	K^*
Anth.gehalt in der Blüte	6	0	H800ATT_MAAGC_3	12,125
Anth.gehalt im Blatt	1	79,211	H800ATT_MAAGC_2	9,629
Anth.gehalt in der Lippe	19	8,85	H700CAA_MATCG_8	12,71
Blütengröße	keine signifikanten Marker			
Rispenlänge	2	92,017	H700CGA_MACGT_5	10,695
	4	0	H700CGA_MACAT_7	9,825
	1	22,689	H700CAT_MGACT_1	9,822
	10	24,41	H800ATT_MACAT_8	9,467
loglip	19	8,85	H700CAA_MATCG_8	12,71

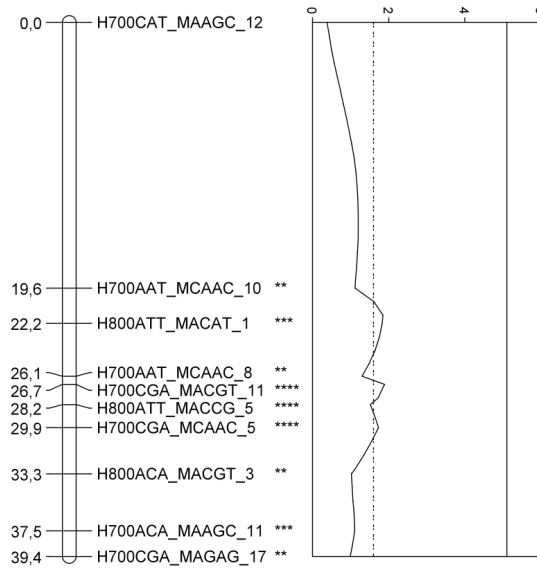


Abbildung 3.8: **QTL-Analyse für Blütengröße:** Kopplungsgruppe 1 der maternalen Karte; Links ist die Kopplungsgruppe mit dem Abstand in cM, dem Markernamen und den Signifikanzen des Kruskal-Wallis-Tests als Sternchen zu sehen. Rechts ist das Intervallmapping als Graph entlang der Kopplungsgruppe (senkrecht) über die LOD-Werte (waagrecht) dargestellt. Die kopplungsgruppen- und genomweite Signifikanzschwelle ist mit einer gestrichelten bzw. durchgezogenen Linie gekennzeichnet (Signifikanzschwellen sind in Tabelle 5.14 aufgeführt).

umliegenden Region ein signifikanter Einfluss festgestellt werden.

Abbildung 3.11 zeigt vier Kopplungsgruppen, die einen Einfluss auf den Anthocyan-gehalt im Blatt haben. Auf Kopplungsgruppe 19 der maternalen Karte ist der Marker H700CGA_MACGA_19 lokalisiert, welcher im Kruskal-Wallis-Test einen signifikanten Einfluss auf das Merkmal zeigt. Die Kopplungsgruppen 2, 9 und 13 zeigen im Intervall-mapping Regionen, die die kopplungsgruppenweite Signifikanzschwelle überschreiten und somit auch einen Einfluss auf das Merkmal haben können.

Abbildung 3.12 und 3.13 zeigen die Kopplungsgruppen die einen Einfluss auf das Merkmal Rispenlänge zeigen. Auf Kopplungsgruppe 18 der maternalen Karte und auf den Kopplungsgruppen 1, 2, 4 und 10 der paternalen Karte sind Marker lokalisiert, die im Kruskal-Wallis-Test einen signifikanten Einfluss zeigen. In Kopplungsgruppe 4 und 13 der maternalen Karte und Kopplungsgruppe 15 der paternalen Karte wird das kopplungsgruppenweite Signifikanzniveau überschritten. Es konnte kein Marker bzw. keine Region gefunden werden, der in beiden Tests einen signifikanten Einfluss

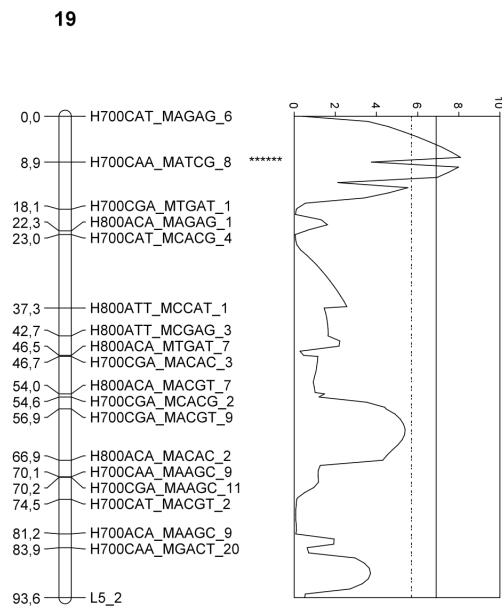
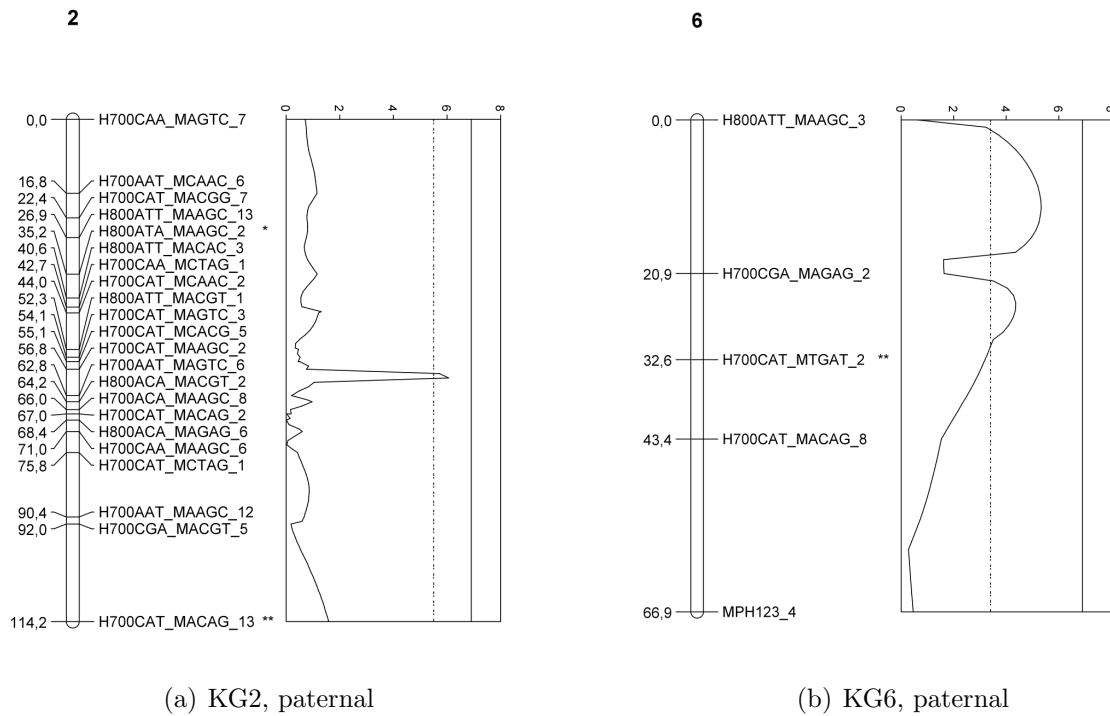


Abbildung 3.9: **QTL-Analyse für Anthocyangehalt in der Lippe:** (a) Kopp-
lungsgruppe 2 der paternalen Karte, (b) Kopplungsgruppe 6 der pa-
ternalen Karte, (c) Kopplungsgruppe 19 der paternalen Karte; Die
Beschreibung entspricht der in Abbildung 3.8.

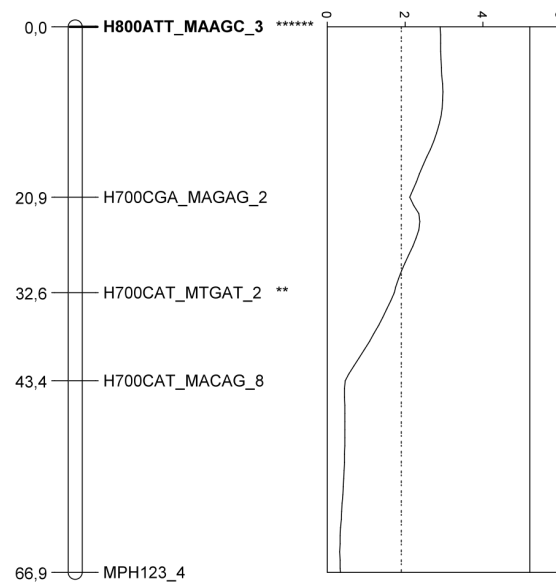
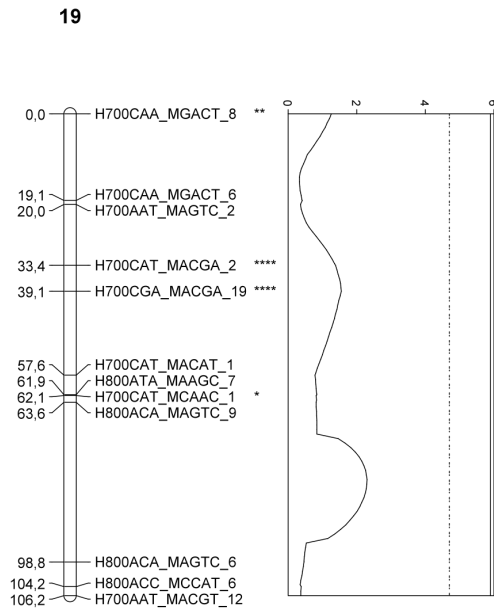
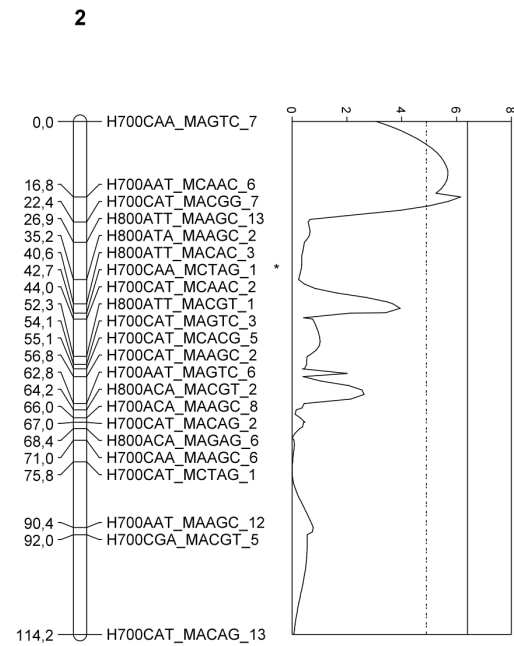


Abbildung 3.10: **QTL für Anthocyangehalt in der Blüte:** Kopplungsgruppe 6 der paternalen Karte; Die Beschreibung entspricht der in Abbildung 3.8.

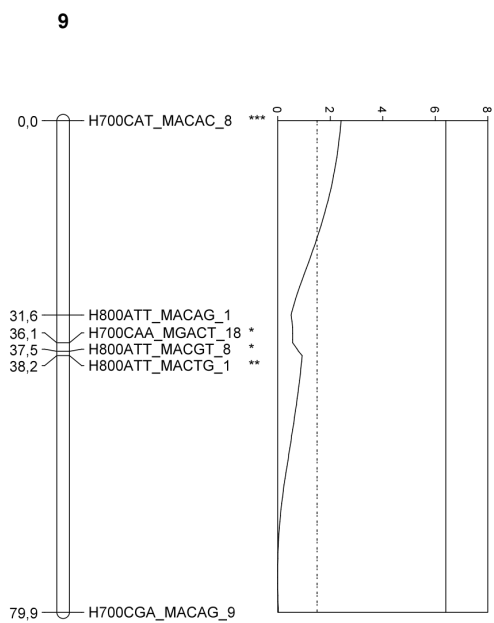
auf das Merkmal zeigten.



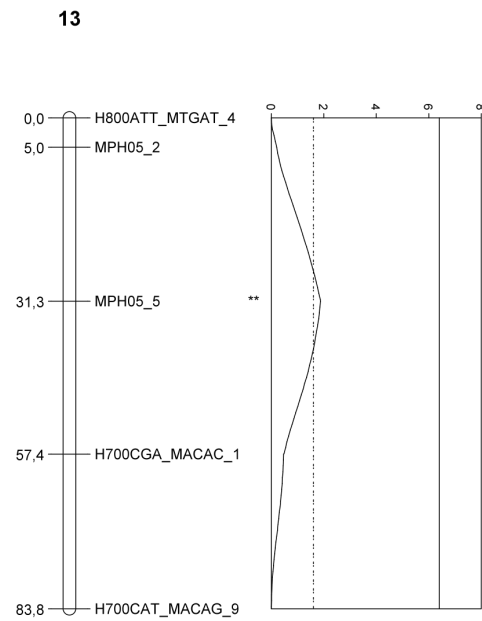
(a) KG19, maternal



(b) KG2, paternal



(c) KG9, paternal



(d) KG13, paternal

Abbildung 3.11: **QTL für Anthocyangehalt im Blatt:** (a) Kopplungsgruppe 19 der maternalen Karte, (b) Kopplungsgruppe 2 der paternalen Karte, (c) Kopplungsgruppe 9 der paternalen Karte, (d) Kopplungsgruppe 13 der paternalen Karte; Die Beschreibung entspricht der in Abbildung 3.8.

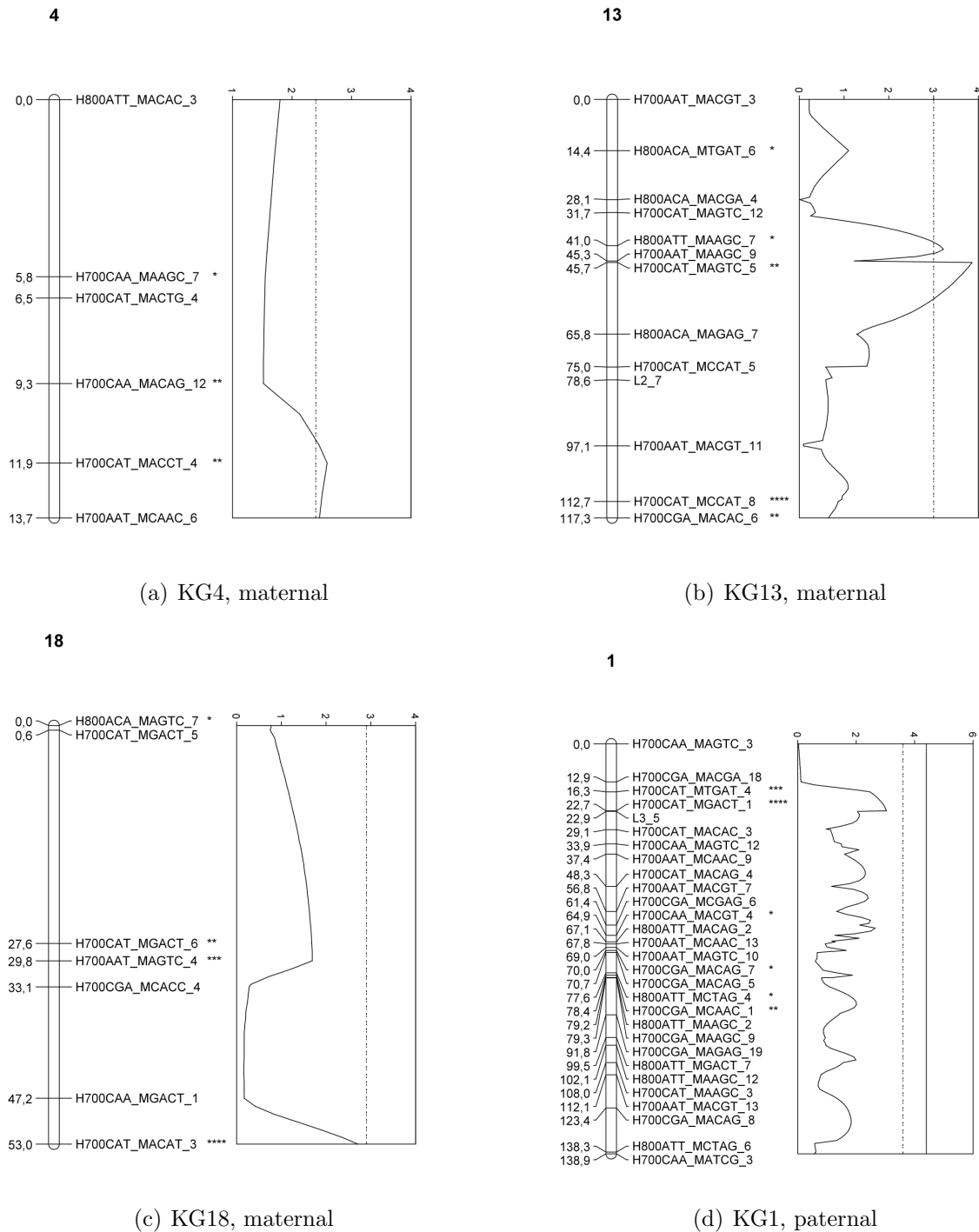
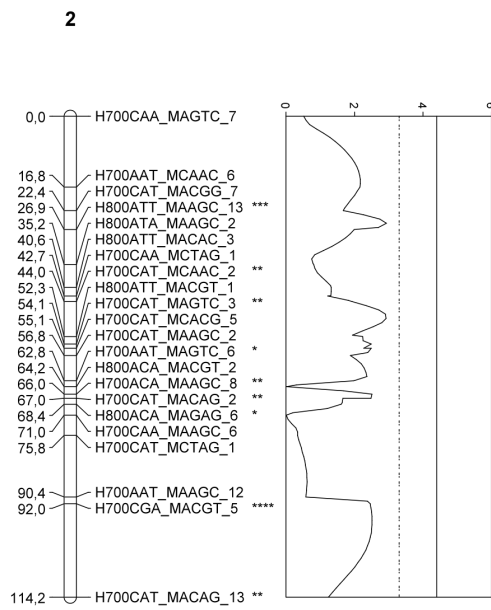
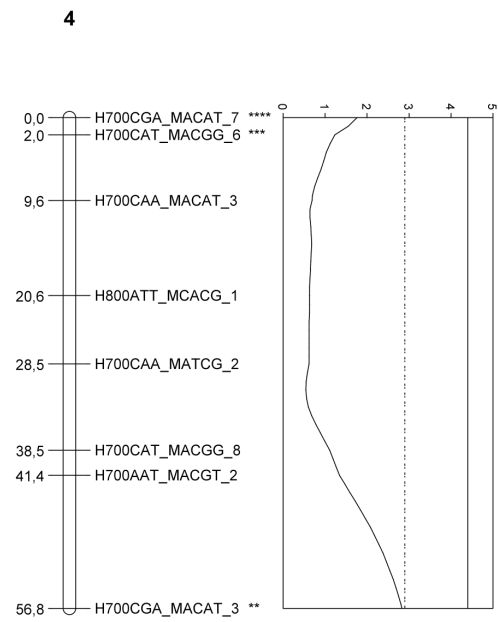


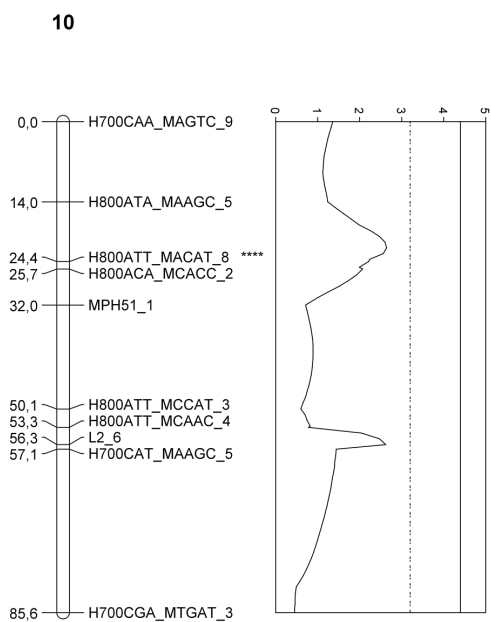
Abbildung 3.12: **QTL für Rispenlänge (Teil 1)**: (a) Kopplungsgruppe 4 der maternalen Karte, (b) Kopplungsgruppe 13 der maternalen Karte, (c) Kopplungsgruppe 18 der maternalen Karte, (d) Kopplungsgruppe 1 der paternalen Karte; Die Beschreibung entspricht der in Abbildung 3.8.



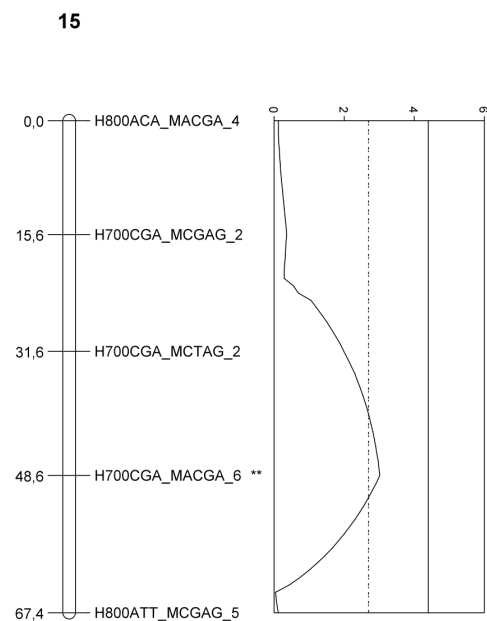
(a) KG2, paternal



(b) KG4, paternal



(c) KG10, paternal



(d) KG15, paternal

Abbildung 3.13: **QTL für Rispenlänge (Teil 2)**: (a) Kopplungsgruppe 2 der paternalen Karte, (b) Kopplungsgruppe 4 der paternalen Karte, (c) Kopplungsgruppe 10 der paternalen Karte, (d) Kopplungsgruppe 15 der paternalen Karte; Die Beschreibung entspricht der in Abbildung 3.8.

3.2 Genomweite Assoziationsstudie

Auf die Genotypensammlung, bestehend aus 144 aktuellen *Phalaenopsis* Hybriden und zehn Wildarten (davon zwei in Wiederholung), wurden 25 AFLP Primerkombinationen angewendet. Pro Primerkombination wurden durchschnittlich 128 Banden pro Genotyp erzeugt, von denen im Durchschnitt 20 Marker ausgewertet werden konnten. Insgesamt wurden 492 AFLP-Marker für die genomweite Assoziationsstudie verwendet.

Bei der Anwendung der Mikrosatellitenmarker auf die Genotypensammlung erzeugten von 22 Mikrosatellitenprimerpaaren sechs ein Bandenmuster, das eindeutig ausgewertet werden konnte. Dies resultierte in 59 dominant auswertbaren Mikrosatellitenmarkern.

Die Ergebnisse der genomweiten Assoziationsstudie werden in den folgenden Kapiteln dargestellt. Zunächst wurde die genetische Diversität zwischen den Genotypen mit klassischen distanzbasierten Methoden berechnet und in einem Dendrogramm dargestellt. Anschließend wurde die Populationsstruktur und das Kopplungsungleichgewicht der zu untersuchenden Population ermittelt. Nachdem die Voraussetzungen für eine genomweite Assoziationsstudie bekannt waren, konnten die Marker-Eigenschafts-Assoziationen mit zwei unterschiedlichen Methoden berechnet werden.

3.2.1 Diversitätsanalyse

3.2.1.1 Variationskoeffizient

Die Diversitätsanalyse der 144 Genotypen wurde mit 492 AFLP und 59 dominant ausgewerteten Mikrosatellitenmarkern durchgeführt.

Um festzustellen, ob ausreichend Marker für eine Diversitätsanalyse zur Verfügung stehen, wurde zunächst der Variationskoeffizient nach Zeid et al. (2003) berechnet. Abbildung 3.14 stellt die Verteilung der Variationskoeffizienten dar. In Abbildung 3.14 (a) sind alle Variationskoeffizienten dargestellt, inklusive der Ausreißer mit einem Variationskoeffizienten von über eins. Der maximale Variationskoeffizient beträgt 1,085 und der minimale Variationskoeffizient 0,0658 (abgesehen von den Vergleichen der identischen Genotypen mit einem Variationskoeffizient von 0). Für eine bessere Darstellung des Hauptanteils der Variationskoeffizienten wurde in Abbildung 3.14 (b) eine andere Skalierung der x-Achse gewählt. Der Hauptteil der Variationskoeffizienten befindet sich zwischen 0,07 und 0,15 mit einem mittleren Variationskoeffizienten von 0,093. Damit befindet sich der Mittelwert der Variationskoeffizienten unter der von Zeid et al. (2003) angegebenen Schwelle von 0,1 und die Anzahl an Markern ist demzufolge ausreichend für eine Diversitätsanalyse mit 146 Individuen basierend auf

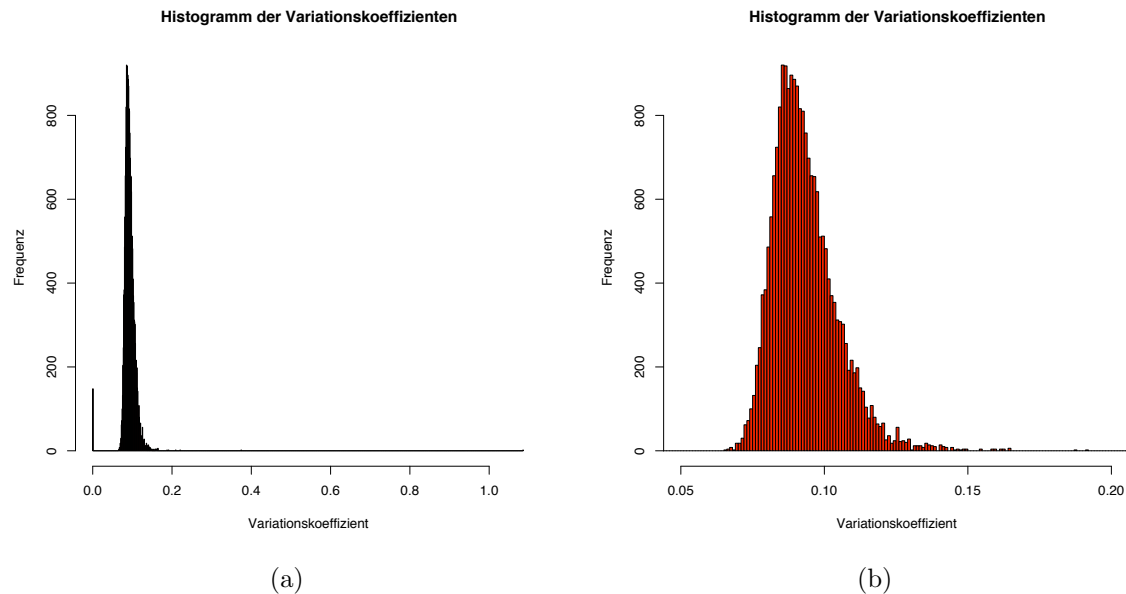


Abbildung 3.14: **Histogramm der Variationskoeffizienten:** (a) Übersicht aller Variationskoeffizienten, (b) Teilansicht; es werden nur die Häufigkeiten der Variationskoeffizienten zwischen 0,05 und 0,20 gezeigt

492 AFLP-Markern.

3.2.1.2 Phylogenetischer Baum

In Abbildung 3.15 ist das Dendrogramm berechnet nach der Neighbour-Joining Methode (Saitou and Nei (1987)) dargestellt. Das Dendrogramm enthält 146 Individuen (=144 Genotypen, zwei Genotypen in Wiederholung [0001 und 0011 = *Ph. violaceae*; 0002 und 0012 = *Ph. hieroglyphica*]) und basiert auf 492 AFLP- und 59 dominanten Mikrosatelliten-Markern. Die Distanzen wurden nach der Methode von Nei and Li (1979) in PHYLIP 3.6.8 berechnet. Die kleinste Distanz wurde zwischen den Genotypen 0819 und 0820 mit 0,002689 detektiert. Die größte Distanz besteht zwischen den Genotypen 0739 und 0001 bzw. 0011 (*Ph. violaceae* in Wiederholung) mit einem Wert von 0,50112. Es konnte keine Gruppierung der Genotypen nach Herkunft festgestellt werden.

Allerdings kann eine Gruppierung der Genotypen in vier Hauptgruppen mit jeweils zwei, drei und null Untergruppen vorgenommen werden. Die Hauptgruppen sind in Abbildung 3.15 mit den Zahlen eins bis vier und die Untergruppen sind mit Buchstaben A bis C gekennzeichnet. Im Anhang (Tabelle 5.15 bis 5.23) sind die Genotypen entsprechend der Reihenfolge im Dendrogramm (siehe Abbildung 3.15) aufgeführt.

In Gruppe 4 sind ausschließlich Genotypen mit weißer Blütenfarben mit höchstens einem Hauch rosa vertreten. In dieser Gruppe fällt vor allem die starke farbliche Ab-

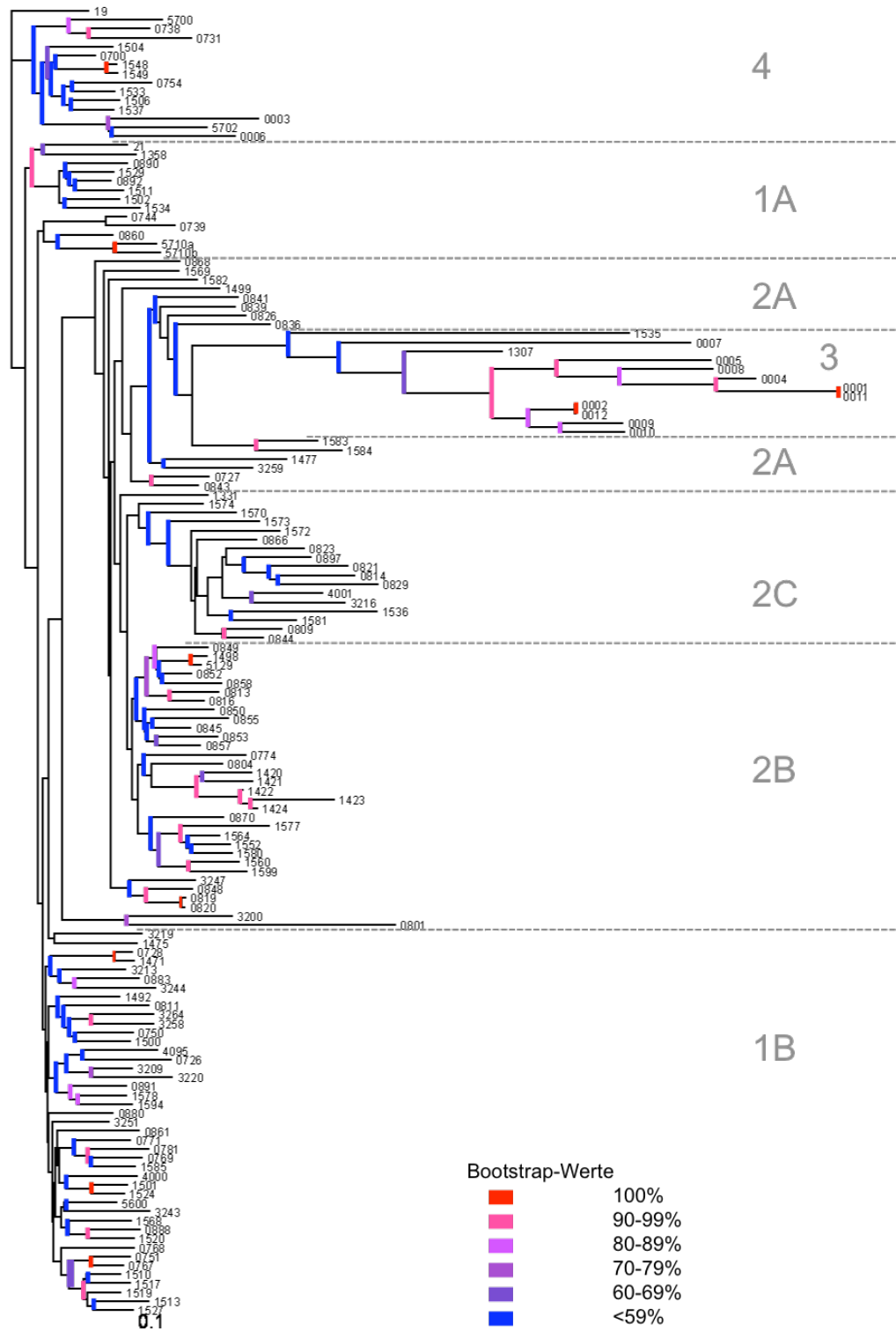


Abbildung 3.15: Neighbour-Joining Dendrogramm; 0001 bis 0012 Wildtypen; Die Verzweigungen sind entsprechend der Bootstrap-Werte farblich markiert. Das Dendrogramm besteht aus 146 Genotypen (zwei davon in Wiederholung) und basiert auf 492 AFLP- und 59 dominanten Mikrosatellitenmarkern.

grenzung der Lippe gegenüber den zumeist reinweißen Tepalen, sowie die abweichende Blütenform auf. Des Weiteren sind in dieser Gruppe die Wildtypen *Ph. philippinensis* (0003) und *Ph. stuartiana* (0006) zu finden.

In Gruppe 1A sind weiße, hellrosa- und pinkfarbene Blüten zu finden, hingegen wird Gruppe 1B hauptsächlich von Genotypen mit weißen Blüten dominiert. Allerdings sind auch rosafarbene Genotypen an den äußeren Rändern der Gruppe 1B gruppiert (z.B. die Genotypen 1471, 1475, 1492, 0811 und 0726). Des Weiteren ist auffällig, daß die Blüten der Genotypen dieser Gruppe alle amabilisförmig und geschlossen sind.

Gruppe 2A ist eine die Gruppe 3 flankierende Gruppe. In Gruppe 3 sind vorwiegend die Wildtypen zu finden, sowie Hybriden deren Elter ein Wildtyp ist (Genotyp 1307 hat *Ph. violaceae* als Elter). Gruppe 2A ist im Phänotyp der Blüte relativ heterogen. Es sind vorwiegend gelbe bzw. gelbliche sowie gemusterte Blüten zu finden.

Die Gruppe 2B besteht fast ausschließlich aus pinkfarbenen bzw. violetten Blütentypen. In Gruppe 2C sind hingegen gelbe, pinkfarbene und stark gemusterte Blüten zu finden. Die Diversität ist in Gruppe 2B und 2C höher als in Gruppe 1B.

3.2.2 Populationsstruktur

3.2.2.1 Anzahl an Subpopulationen

Um die Populationsstruktur zu bestimmen, wurden zunächst vier unabhängige Durchläufe in STRUCTURE 2.2 gerechnet, in denen jeweils 2-15 Subpopulationen (K) angenommen wurden. Dabei wurden die untersuchten Genotypen jeweils in die angenommene Anzahl von Subpopulationen eingeteilt. Anschließend wurde aus allen Durchläufen die vermutlich 'wahre' Anzahl der Subpopulationen bestimmt.

In den folgenden Kapiteln wird deutlich werden, dass die untersuchte *Phalaenopsis* Population aus 10-12 Subpopulationen besteht. Deshalb sollen hier nur die so genannten 'Bar Plots' aus STRUCTURE 2.2 mit K=10 bis K=12 gezeigt werden. Dabei stellen die einzelnen 'bars' Säulen dar, die die Zugehörigkeit einzelner Genotypen an den einzelnen Subpopulationen farblich zeigen. Eine Säule stellt dabei 100% dar und der jeweilige Anteil einer Farbe zeigt den Anteil an der entsprechenden Subpopulation in diesem Genotyp.

Die Abbildungen 3.16 bis 3.18 zeigen die genetische Zuordnung der einzelnen Genotypen zu den jeweiligen Subpopulationen. Jede Farbe stellt eine Subpopulation dar, wobei die Farben je Bild neu für jede Subpopulation zugeordnet wurden und deshalb nicht über die einzelnen Bilder hinweg verglichen werden können. Das Programm STRUCTURE 2.2 kodiert die Genotypen numerisch. Welcher Genotyp hinter den Zahlen steht, ist dem Anhang (Tabelle 5.15 bis 5.23) zu entnehmen. Beispielhaft soll die Graphik anhand der Genotypen 1307 und *Ph. violaceae* erläutert werden. Der Ge-

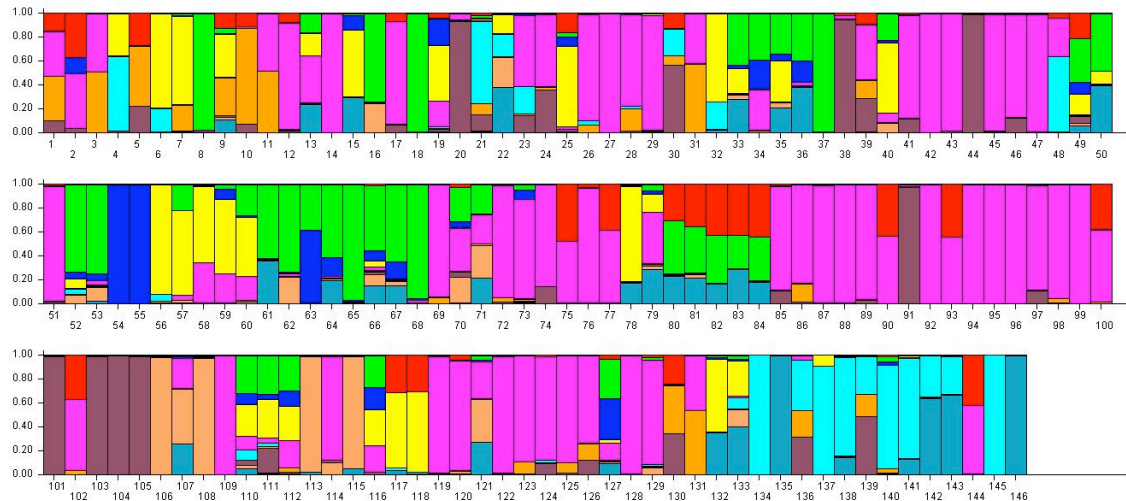


Abbildung 3.16: **Populationsstruktur mit K=10 Subpopulationen:** Populationsstruktur von 146 Individuen einer repräsentativen Sammlung von *Phalaenopsis* Hybriden. Jedes Individuum ist durch einen in verschiedene Farben unterteilten Balken dargestellt. Jede Farbe repräsentiert eine Subpopulation und dementsprechend ist die Zugehörigkeit eines Individuums zu einer Subpopulation durch den Anteil der entsprechenden Farbe dargestellt. Die Bezeichnung der Genotypen ist dem Anhang zu entnehmen

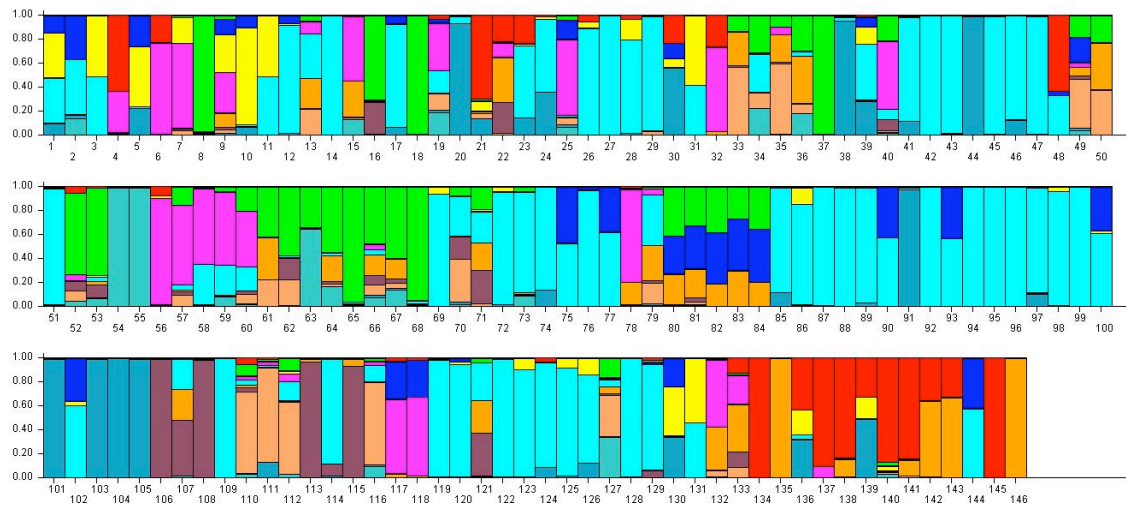


Abbildung 3.17: **Populationsstruktur mit K=11 Subpopulationen:** Populationsstruktur von 146 Individuen einer repräsentativen Sammlung von *Phalaenopsis* Hybriden. Jedes Individuum ist durch einen in verschiedene Farben unterteilten Balken dargestellt. Jede Farbe repräsentiert eine Subpopulation und dementsprechend ist die Zugehörigkeit eines Individuums zu einer Subpopulation durch den Anteil der entsprechenden Farbe dargestellt. Die Bezeichnung der Genotypen ist dem Anhang zu entnehmen

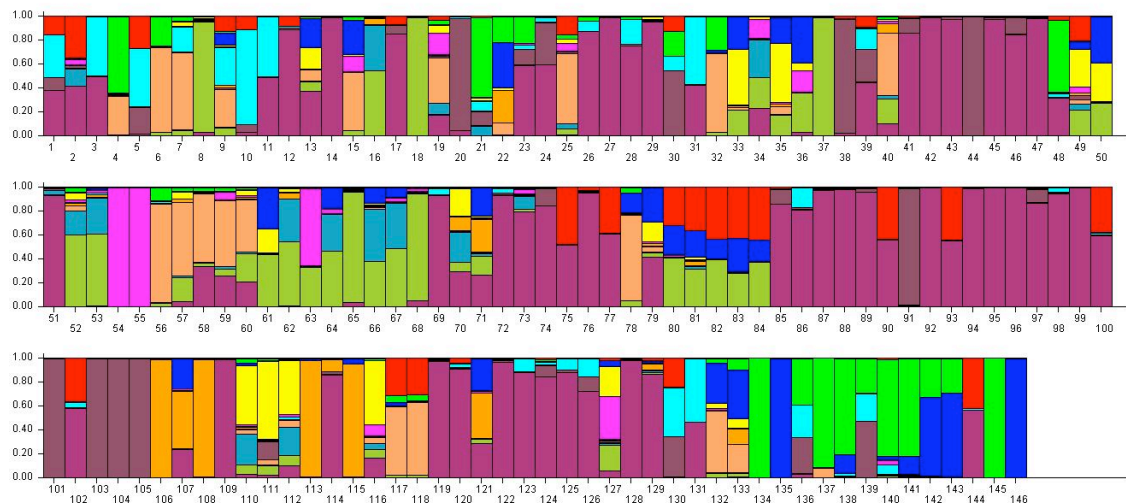


Abbildung 3.18: **Populationsstruktur mit K=12 Subpopulationen:** Populationsstruktur von 146 Individuen einer repräsentativen Sammlung von *Phalaenopsis* Hybriden. Jedes Individuum ist durch einen in verschiedene Farben unterteilten Balken dargestellt. Jede Farbe repräsentiert eine Subpopulation und dementsprechend ist die Zugehörigkeit eines Individuums zu einer Subpopulation durch den Anteil der entsprechenden Farbe dargestellt. Die Bezeichnung der Genotypen ist dem Anhang zu entnehmen

notyp *Ph. violaceae* ist mit der Nummer 134 und 145 (eine Wiederholung) versehen. In Bild 3.16 ist dieser Genotyp nahezu komplett hellblau. Der Genotyp 1307 ist mit der Nummer 4 versehen und zeigt in Bild 3.16 zu ca. 60% die Farbe hellblau. Dies bedeutet, dass dieser Genotyp zu 60% der gleichen Subpopulation wie *Ph. violaceae* zuzuordnen ist. Da *Ph. violaceae* ein Elter des Genotypen 1307 ist, zeigt sich hier dass die Zuordnung der Genotypen zu einzelnen Subpopulationen nachvollziehbar ist und aufgrund genetischer Ähnlichkeiten erfolgt.

In Abb 3.19 sind die Populationsstrukturen gegenüber dem Dendrogramm dargestellt. Es ist zu sehen, dass die Populationsstrukturen in etwa der Anordnung im Dendrogramm entsprechen.

3.2.2.2 Ermittlung der Anzahl an Subpopulationen nach Pritchard et al. (2000)

Die Berechnung der Anzahl an Subpopulationen wurde entsprechend der Methodik, beschrieben in Pritchard et al. (2000) sowie im Handbuch der Software STRUCTURE 2.2, durchgeführt. In Tabelle 3.8 sind die Ergebnisse dieser Berechnungen für die jeweiligen Durchläufe zu finden.

Die Ermittlung der Anzahl der Subpopulationen basiert auf einem Vergleich, der für

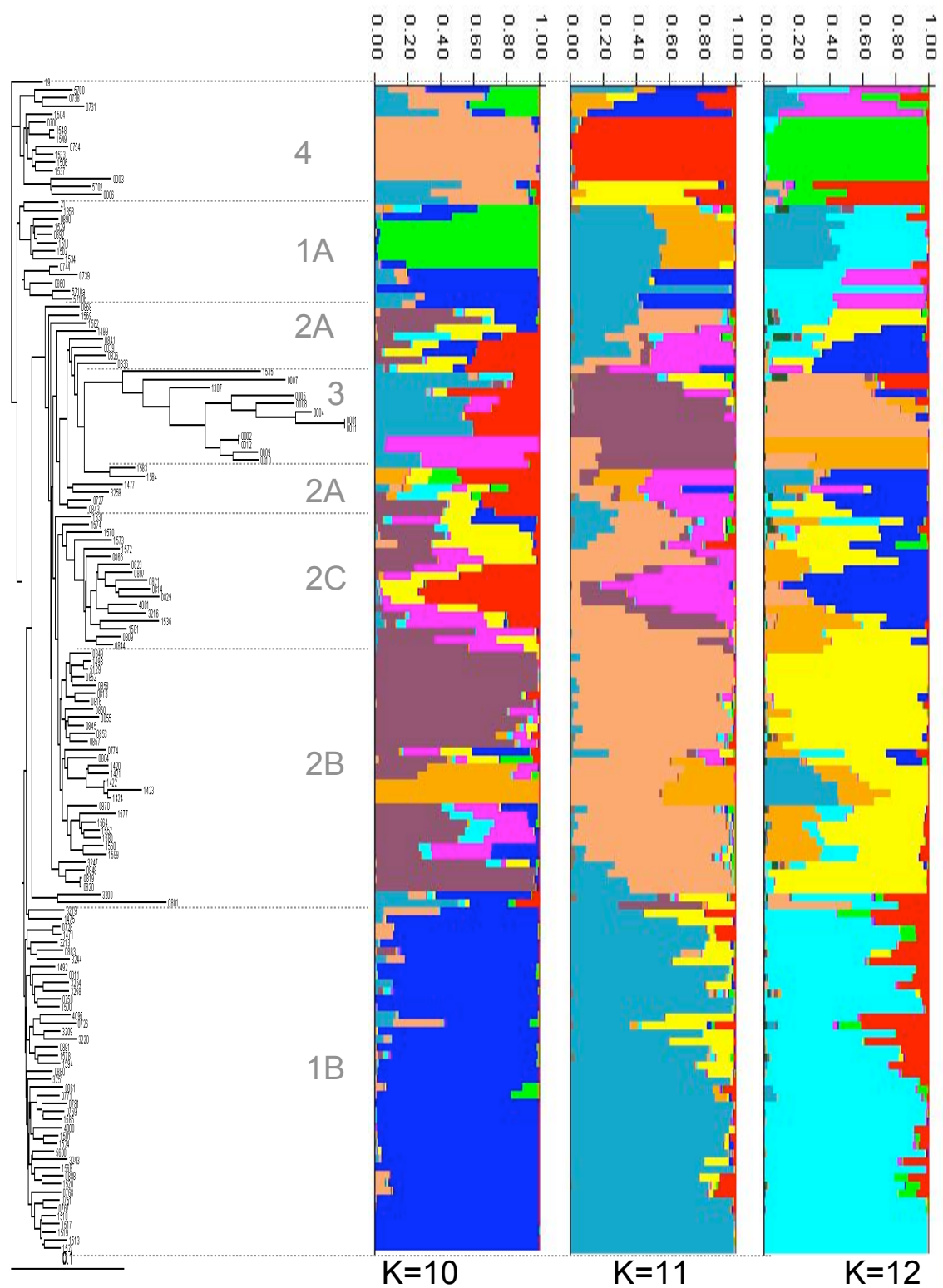


Abbildung 3.19: Gegenüberstellung des Dendrogramms mit den einzelnen Subpopulationen

jede angenommene Anzahl von Subpopulationen (K) von STRUCTURE 2.2 ausgegebenen geschätzten Wahrscheinlichkeiten (lnP(D)-Werte). Die Anzahl an Subpopulationen ist dort zu wählen, wo der lnP(D)-Wert sich im Plateau befindet. Dies ist dann gegeben, wenn die Berechnung gleich eins ist. In Tabelle 3.8 sind diese Werte grau unterlegt. In der vorliegenden Arbeit wurde diese Berechnung für alle vier Durchläufe separat durchgeführt. Für den ersten und dritten Durchlauf ist die Anzahl der Subpopulationen jeweils bei 10 zu wählen. Im zweiten Durchlauf ist das Plateau bei einer Anzahl von 12 und im vierten Durchlauf bei 8 Subpopulationen erreicht.

Tabelle 3.8: Ermittlung der Anzahl der Subpopulationen nach Pritchard et al. (2000): ermittelte 'geschätzte posterior Wahrscheinlichkeiten' für K=2-15 und vier Durchläufe

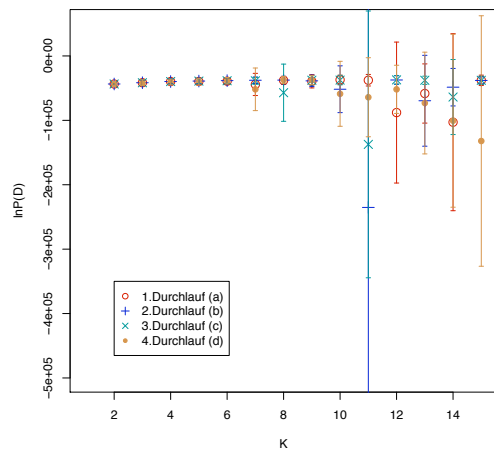
Anzahl an Subpopulationen (K)	1. Durchlauf	2.Durchlauf	3.Durchlauf	4.Durchlauf
2	0	0	0	0
3	0	0	0	0
4	0	0	0	0
5	0	0	0	0
6	0	0	0	0
7	0	7,00E-188	0	0
8	0	8,00E-84	0	1
9	0	0	1,00E-90	5,00E-98
10	1	0	1	0
11	3,00E-284	0	0	0
12	0	1	8,12E+157	0
13	0	0	2,00E-118	0
14	0	0	0	0
15	0	1,00E-305	1,00E-128	0

3.2.2.3 Ermittlung der Anzahl an Subpopulationen nach Cockram et al. (2008)

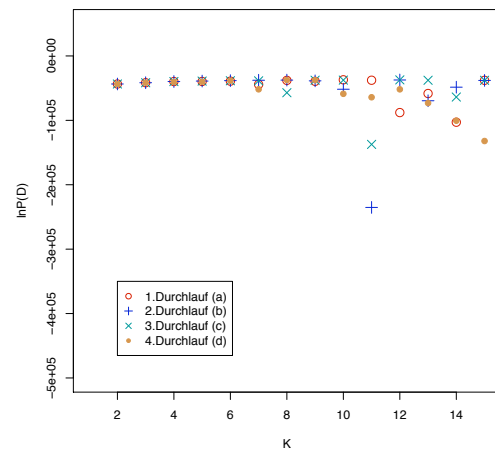
Bei dieser Methode wird die Anzahl der Subpopulationen zunächst graphisch geschätzt und anschließend durch einen Vergleich der Q-Matrizen verifiziert (siehe Abschnitt 2.5.6).

Graphische Schätzung der Anzahl der Subpopulation

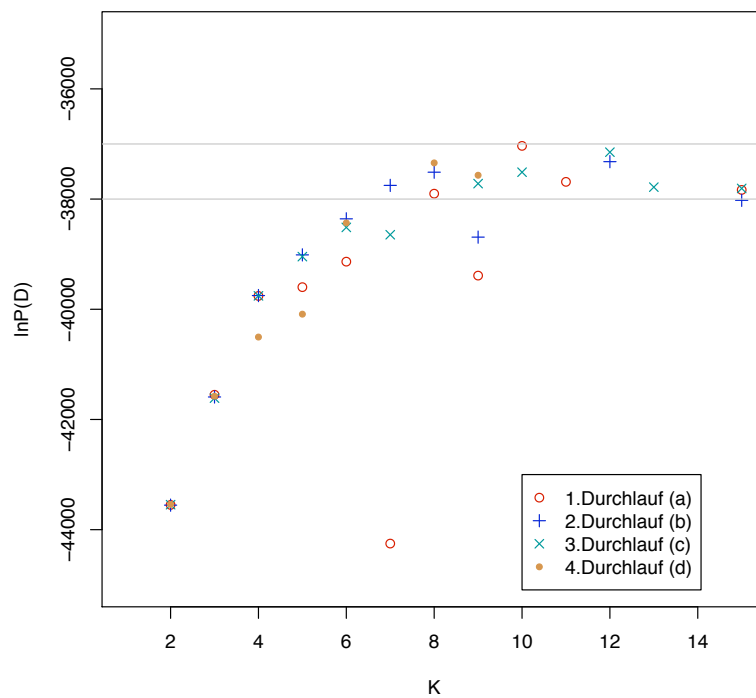
In Abbildung 3.20 sind die lnP(D)-Werte in Abhängigkeit von der geschätzten Anzahl der Subpopulation (K) aufgetragen. Zunächst ist die Gesamtübersicht der berechneten lnP(D)-Werte mit (a) und zur besseren Übersicht ohne (b) Varianzen dargestellt. Zwischen den Wiederholungen der Berechnungen für jedes K sind bis K=6 kaum Abweichungen zu sehen. Diese Abweichungen, sowie die Standardabweichung



(a)



(b)



(c)

Abbildung 3.20: Graphische Darstellung der $\ln P(D)$ -Werte aus STRUCTURE 2.2 in Abhängigkeit der Anzahl an Subpopulationen (K). (a) Gesamtübersicht mit Varianzen; (b) Gesamtübersicht ohne Varianzen; (c) Teilübersicht zur Bestimmung des Plateaus. Das Plateau liegt zwischen -38000 und -37000 und ist durch zwei graue Linien begrenzt.

der einzelnen Berechnung, werden mit zunehmender Anzahl an geschätzten Subpopulationen (K) größer. Um ein mögliches Plateau der $\ln P(D)$ -Werte besser identifizieren zu können, wurde eine feinere Skalierung der y-Achse vorgenommen und ist in Abbildung 3.20 (c) dargestellt. Diese Graphik zeigt ein Plateau der $\ln P(D)$ Werte zwischen -38 000 und -37 000. Die zugehörigen K Werte liegen zwischen 7 und 12. Demzufolge ist die Anzahl der Subpopulationen zwischen $K=7$ bis $K=12$ zu wählen.

Vergleich der Q-Matrizen

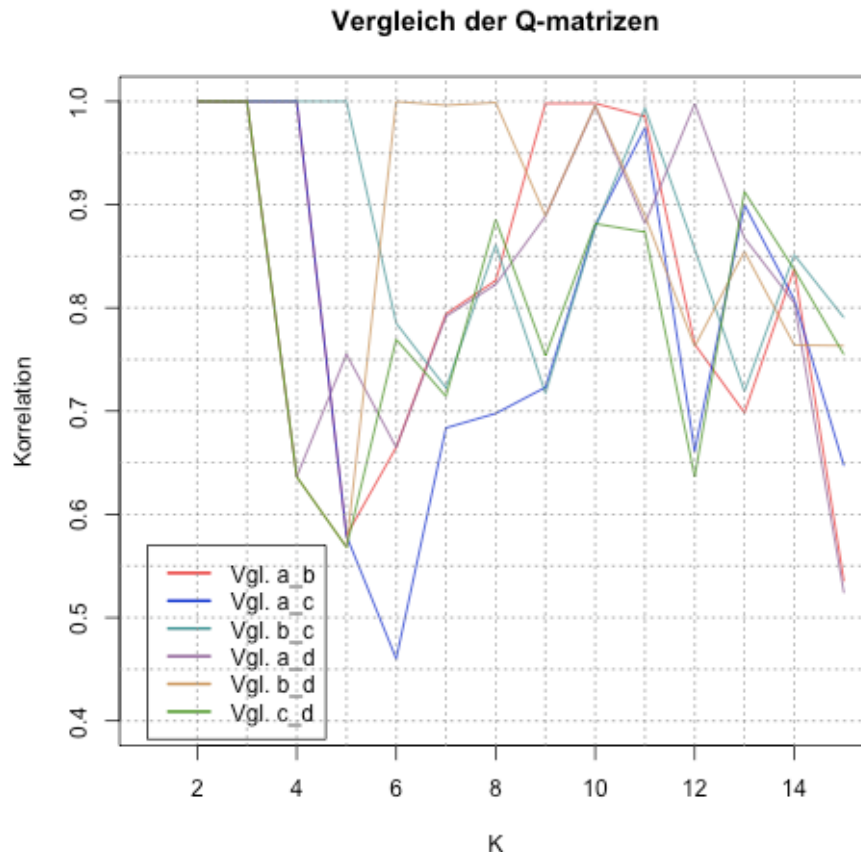


Abbildung 3.21: **Graphische Darstellung der Korrelation zwischen den Q Matrizen:** die Vergleiche der jeweiligen Durchläufe (a-d) sind jeweils farblich unterschieden. Die beste Korrelation ist bei der Anzahl an Subpopulationen (K) zu wählen, wo alle Vergleiche eine möglichst gute Korrelation zeigen. Hier ist dies bei $K=10$ und $K=11$ gegeben.

In Abbildung 3.21 sind die Vergleiche der Q-Matrizen über jede Wiederholung in Abhängigkeit der geschätzten Anzahl der Subpopulation (K) aufgetragen. Es wurden immer die Q-Matrizen zweier Durchläufe miteinander verglichen. Es ist zu sehen, daß die Korrelationen der Q-Matrizen bei $K=10$ und bei $K=11$ am höchsten sind. Hier sinkt der mittlere maximale Korrelationswert bei keinem Vergleich unter 0,85.

Bei $K=4$ bis 9 und $K=12$ bis 15 sind Vergleiche zu sehen, deren mittlerer maximaler Korrelationswert rapide absinkt. Nur bei $K=2$ und 3 ist der mittlere maximale Korrelationswert maximal.

Ermittlung der Anzahl der Subpopulationen

Nach J. Cockram (persönliche Mitteilung) ist die Anzahl der Subpopulationen für die zu untersuchende Population dort zu wählen, wo der $\ln P(D)$ Wert im Plateau liegt und der Vergleich der entsprechenden Q-Matrizen eine gute bis sehr gute Korrelation zeigt.

In der vorliegenden Studie ist die Anzahl der Subpopulationen bei 10 oder 11 zu legen, da hier die besten Korrelationen (größer als 0,85) erreicht werden und sich deren $\ln P(D)$ -Werte im Plateau befinden. Für $K=10$ spricht, dass hier das Plateau mit einer Steigung von nahezu null erreicht ist. Hier zeigt der Vergleich zwischen den Q-Matrizen einer sehr gute Korrelation zwischen dem ersten, zweiten und vierten Durchlauf und eine ausreichend gute Korrelation zum dritten Durchlauf. Für $K=11$ spricht, die ebenfalls gute Korrelation zwischen allen Durchläufen. Des Weiteren befindet sich der $\ln P(D)$ Wert bei $K=11$ in der maximalen Plateauregion. Auch $K=9$ und $K=12$ sind noch in der maximalen Plateauregion zu finden. Gegen die Anzahl von 9 bzw. 12 Subpopulationen spricht jedoch die geringe Korrelation zwischen den Q-Matrizen dieser Durchläufe. Damit ist die Zuverlässigkeit dieser Berechnungen nicht mehr ausreichend gegeben und von einer Verwendung dieser Anzahl an Subpopulationen ist abzusehen. Der mittlere maximale Korrelationswert ist bei $K=2$ und $K=3$ am größten, jedoch kommen diese Anzahl an Subpopulationen für die untersuchte Population nicht infrage, da sich die $\ln P(D)$ -Werte hier noch nicht im Plateau befinden.

3.2.2.4 Ermittlung der Anzahl an Subpopulationen nach Evanno et al. (2005)

In Abbildung 5.2 ist die Berechnung der Anzahl der Subpopulation nach Evanno et al. (2005) dargestellt. In der ersten Spalte ist der Mittelwert $L(K)$ der $\ln P(D)$ Werte über jedes K in vier Durchläufen dargestellt. Spalte zwei zeigt die Rate der Veränderung der Likelihood Verteilung, berechnet als $L(K)=L(K)-L(K-1)$. Die absoluten Werte der zweiten Ordnung der Rate der Veränderung der Likelihood Verteilung, berechnet als $|L(K)|=|L(K+1)-L(K)|$, ist in der dritten Spalte aufgeführt. Die vierte Spalte zeigt $\Delta K=m|L(K)|/s[L(K)]$. Da hier die Anzahl an Subpopulationen zu finden ist, sind diese Graphiken nochmals in Abbildung 3.22 zur besseren Übersicht dargestellt. Der modale Wert der Verteilung (der höchste ΔK -Wert, sichtbar als Peak im vierten Diagramm) zeigt den vermutlich wahren K -Wert und damit die Anzahl an Subpopulationen. Während im ersten Durchlauf $K=11$ Subpopulationen zu finden sind, sind in den anderen drei Durchläufen die Anzahl von Subpopulationen bei $K=12$ zu finden.

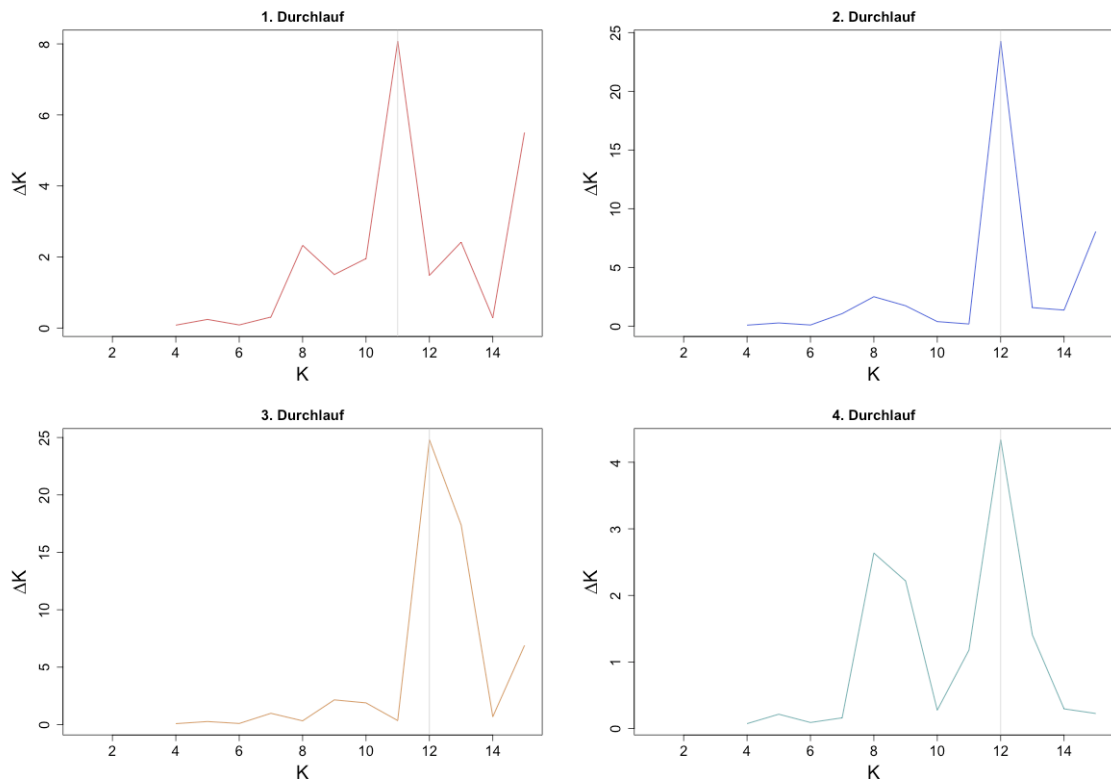


Abbildung 3.22: **graphische Darstellung der Ermittlung der Anzahl der Subpopulationen nach Evanno et al. (2005):** Die vier unabhängigen Durchläufe aus STRUCTURE 2.2 sind in unterschiedlichen Farben dargestellt (1. bis 4. Durchlauf). Für jeden Durchlauf wurde die mögliche Anzahl an Subpopulationen gerechnet. Die entsprechenden Formeln sind dem Text zu entnehmen. Die wahrscheinlichste Anzahl an Subpopulationen ist jeweils bei dem K zu wählen das den höchsten Peak (mit vertikaler grauer Linie unterlegt) zeigt.

Bei allen drei angewendeten Methoden konnte die Anzahl der Subpopulationen und damit die Populationsstruktur, nicht eindeutig bestimmt werden. In der ersten Methode liegt die Anzahl der Subpopulationen vermutlich bei $K=10$. In der zweiten Methode konnten $K=10$ und $K=11$ als mögliche Anzahl an Subpopulation ermittelt werden. In der dritten Methode liegt die Anzahl der Subpopulationen vermutlich bei $K=12$.

Da die Bestimmung der Anzahl der Subpopulation zu keinem eindeutigen Ergebnis geführt hat, wurde die Assoziationsstudie unter der Annahme der verschiedenen möglichen Populationstrukturen durchgeführt.

3.2.3 Kopplungsungleichgewicht zwischen den Markern

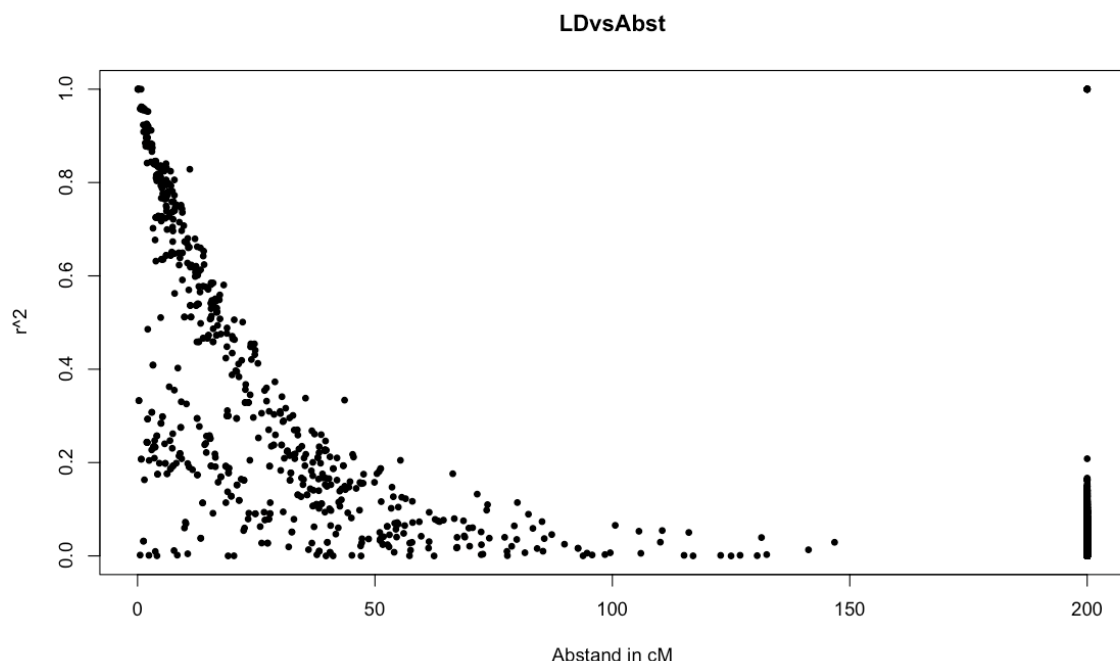


Abbildung 3.23: **Kopplungsungleichgewicht (LD) in Abhängigkeit vom Abstand in cM**; Der Abfall des Kopplungsungleichgewichts wurde in Abhängigkeit zum genetischen Abstand zwischen den genetischen Markern in der Kartierungspopulation '21x19' bestimmt. Es zeigt sich ein Abfall um die Hälfte innerhalb von ca. 30 cM. Das LD zwischen Markern, die nicht auf einer Kopplungsgruppe lokalisiert sind, ist bei 200 cM aufgetragen.

In Abbildung 3.23 ist das Kopplungsungleichgewicht (r^2) zwischen den kartierten Markern in Abhängigkeit zum jeweiligen genetischen Abstand in cM dargestellt. Das Kopplungsungleichgewicht zwischen Markern, die nicht auf einer Kopplungsgruppe

liegen, wurde bei 200 cM dargestellt. Innerhalb von ca. 30 cM sinkt das Kopplungsungleichgewicht um die Hälfte von 1 auf 0,5 ab. Liegen die Marker nicht auf einer Kopplungsgruppe, so ist das Kopplungsungleichgewicht bei fast allen Markern nicht größer als 0,2. Einige nicht gekoppelte Marker zeigen ein hohes Kopplungsungleichgewicht. Hierbei handelt es sich jedoch um Marker die sowohl auf der maternalen als auch auf der paternalen Karte lokalisiert sind (Ankermarker) (siehe Tabelle 3.6).

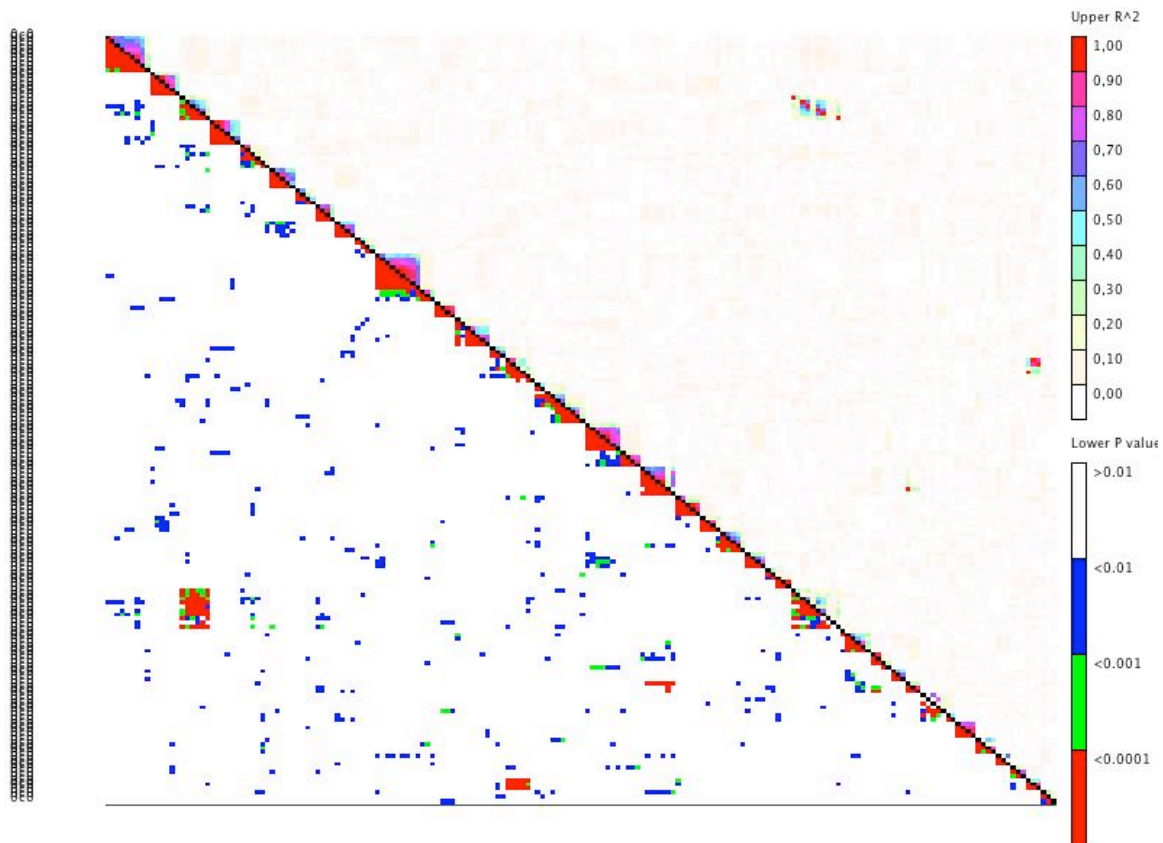


Abbildung 3.24: Kopplungsungleichgewicht zwischen kartierten Markern berechnet und dargestellt mit TASSEL 2.1; oberhalb der Diagonalen ist das Kopplungsungleichgewicht (r^2), und unterhalb der p-Wert dargestellt (nähere Erläuterungen siehe Text)

Eine andere Darstellung des Kopplungsungleichgewichts der kartierten Marker ist in Abbildung 3.24 gezeigt. Diese Abbildung wurde mit dem Programm TASSEL 2.1 erstellt. Oberhalb der Diagonalen ist das Kopplungsungleichgewicht zwischen den Markern in verschiedenen Farbabstufungen zu sehen. Die Farbskala reicht von rot ($r^2=1$) bis hin zu weiß ($r^2=0$). An der Diagonalen gespiegelt, befindet sich in der unteren linken Hälfte der Graphik der jeweilige p-Wert für die Berechnung des Kopplungsungleichgewichts.

Da die Marker in der Reihenfolge der entsprechenden Kopplungsgruppen aufgetragen wurden, befindet sich entlang der Diagonalen jeweils der Vergleich der einzelnen

Marker innerhalb einer Kopplungsgruppe. Wie zu erwarten, zeigen Marker, die sich auf einer Kopplungsgruppe befinden ein hohes Kopplungsungleichgewicht mit einer hohen Signifikanz. Es sind jedoch auch Nester mit hohen Werten für das Kopplungsungleichgewicht außerhalb der Kopplungsgruppen zu finden. Hierbei handelt es sich jedoch um einen Vergleich von Markern, die sowohl auf der maternalen als auch auf der paternalen Kopplungskarte lokalisiert sind (siehe Tabelle 3.6). Da das Kopplungsungleichgewicht hier also zwischen identischen Markern berechnet wird, handelt es sich bei den entsprechenden Werten um Artefakte, die keine weitere Beachtung finden sollen.

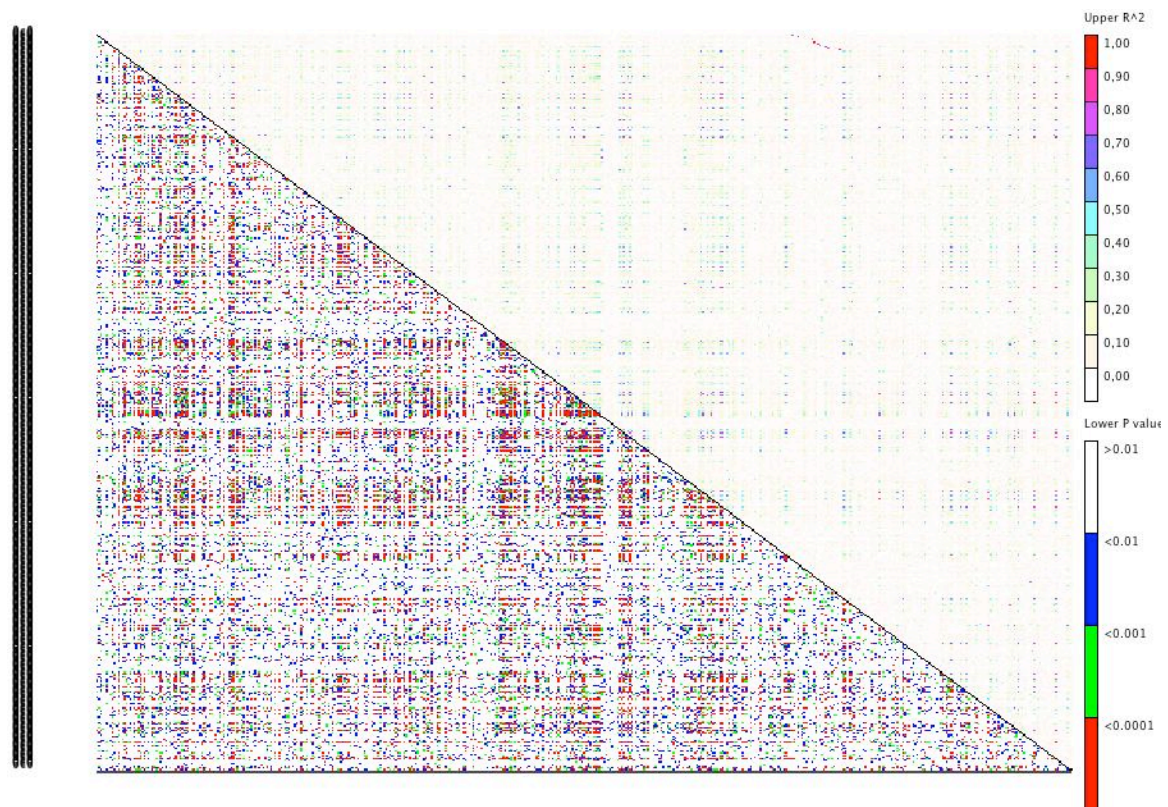


Abbildung 3.25: Kopplungsungleichgewicht zwischen AFLP-Markern berechnet und dargestellt mit TASSEL 2.1; oberhalb der Diagonalen ist das Kopplungsungleichgewicht (r^2), und unterhalb der p-Wert dargestellt (nähere Erläuterungen siehe Text)

Abbildung 3.25 und 3.26 zeigen das Kopplungsungleichgewicht zwischen den Markern, die für die Assoziationsstudie verwendet wurden. Abbildung 3.25 zeigt das Kopplungsungleichgewicht zwischen den AFLP-Markern und Abbildung 3.26 zeigt das Kopplungsungleichgewicht zwischen den Mikrosatellitenmarkern. Oberhalb der Diagonalen ist jeweils das Kopplungsungleichgewicht als r^2 angegeben. Unterhalb der Diagonalen ist deren Signifikanz in unterschiedlichen Stufen als p-Wert dargestellt. Bei beiden Abbildungen ist zu sehen, dass das Kopplungsungleichgewicht zwischen den

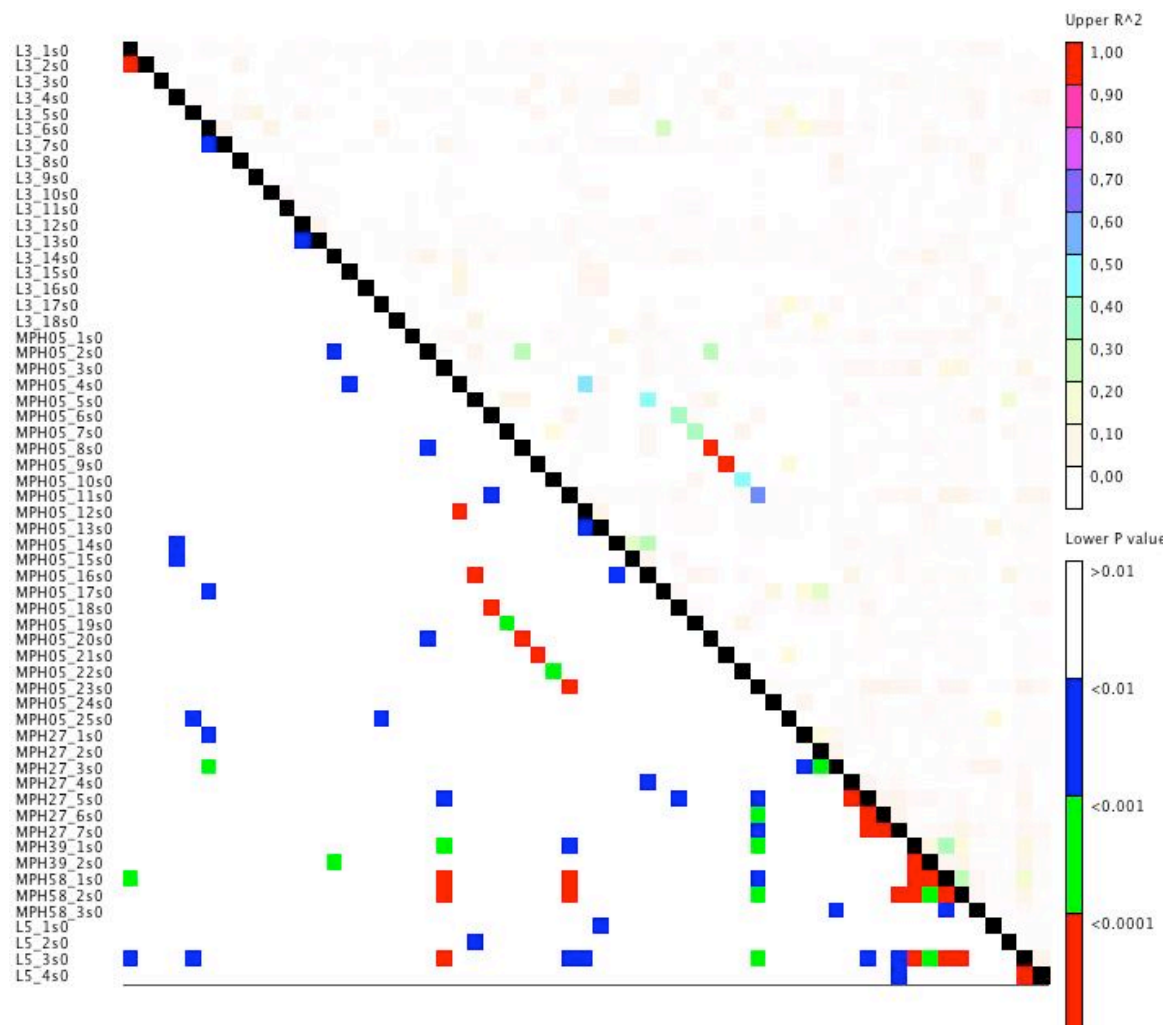


Abbildung 3.26: Kopplungsungleichgewicht zwischen den Mikrosatellitenmarkern berechnet und dargestellt mit TASSEL 2.1; oberhalb der Diagonalen ist das Kopplungsungleichgewicht (r^2), und unterhalb der p-Wert dargestellt (nähere Erläuterungen siehe Text)

Markern relativ klein ist. Die Signifikanzen scheinen bei der Berechnung des Kopplungsungleichgewichts zwischen den AFLP-Markern jedoch höher zu sein als bei der Berechnung des Kopplungsungleichgewichts zwischen den Mikrosatellitenmarkern. Es befinden sich nur ca. 0,9% der AFLP Marker und ca. 0,35% der Mikrosatellitenmarker im Kopplungsungleichgewicht ($r^2 < 0,5$).

3.2.4 Berechnungen der Marker-Eigenschafts-Assoziationen

Aufgrund der mit STRUCTURE 2.2 berechneten Werte, wurde die Assoziationsstudie mit $K=10$, $K=11$ und $K=12$ durchgeführt.

3.2.4.1 Assoziationsstudie mit R 2.8.1

Die Assoziationsstudie in R 2.8.1 wurde auf Grundlage von 492 AFLP-Markern durchgeführt. Dementsprechend wurde die Signifikanzschwelle folgendermaßen berechnet:

$$p = -\text{Log}_{10}(0,05/492) = 3,99 \quad (3.1)$$

In Abbildung 3.27 ist beispielhaft die Ausgabe der Berechnung mit R 2.8.1 für die Eigenschaft Blütengröße und der Anzahl an Subpopulationen $K=10$ aufgeführt.

Die folgende Beschreibung trifft auf alle anderen Berechnungen gleichermaßen zu. Die Graphen links (A und C) zeigen die Berechnung der Assoziationen mit der Populationsstruktur als Kovariate. Die Graphen rechts (B und D) zeigen die Berechnung ohne die Populationsstruktur. Oben (A und B) sind die Graphen zu sehen, die ohne und unten (C und D) sind die Graphen zu sehen, die mit der 'Genomic Control' (GC) Korrektur berechnet wurden. Das Einbeziehen der Populationsstruktur zeigt eine deutliche Verringerung der Marker mit signifikanten p-Werten (vgl. B und D mit A und C). Die 'Genomic Control' Korrektur ergibt eine allgemeine Verringerung der p-Werte (vgl. A und B mit C und D). Das bonferroniadjustierte Signifikanzniveau ist bei 3,99 als horizontale schwarze Linie dargestellt. Bei einem Vergleich der Graphen B und D fällt auf, dass in dem gezeigten Beispiel durch die GC-Korrektur alle berechneten p-Werte der Marker unterhalb des Signifikanzniveaus liegen.

Die Graphen '(C) S.Association.GC.P' aller anderen Berechnungen ist dem Anhang 5.7 zu entnehmen. In Tabelle 3.9 ist zusammenfassend die Anzahl der Marker aufgeführt, die nach den Berechnungen mit R 2.8.1, Assoziationen zu der jeweiligen Eigenschaft zeigen.

Blütengröße

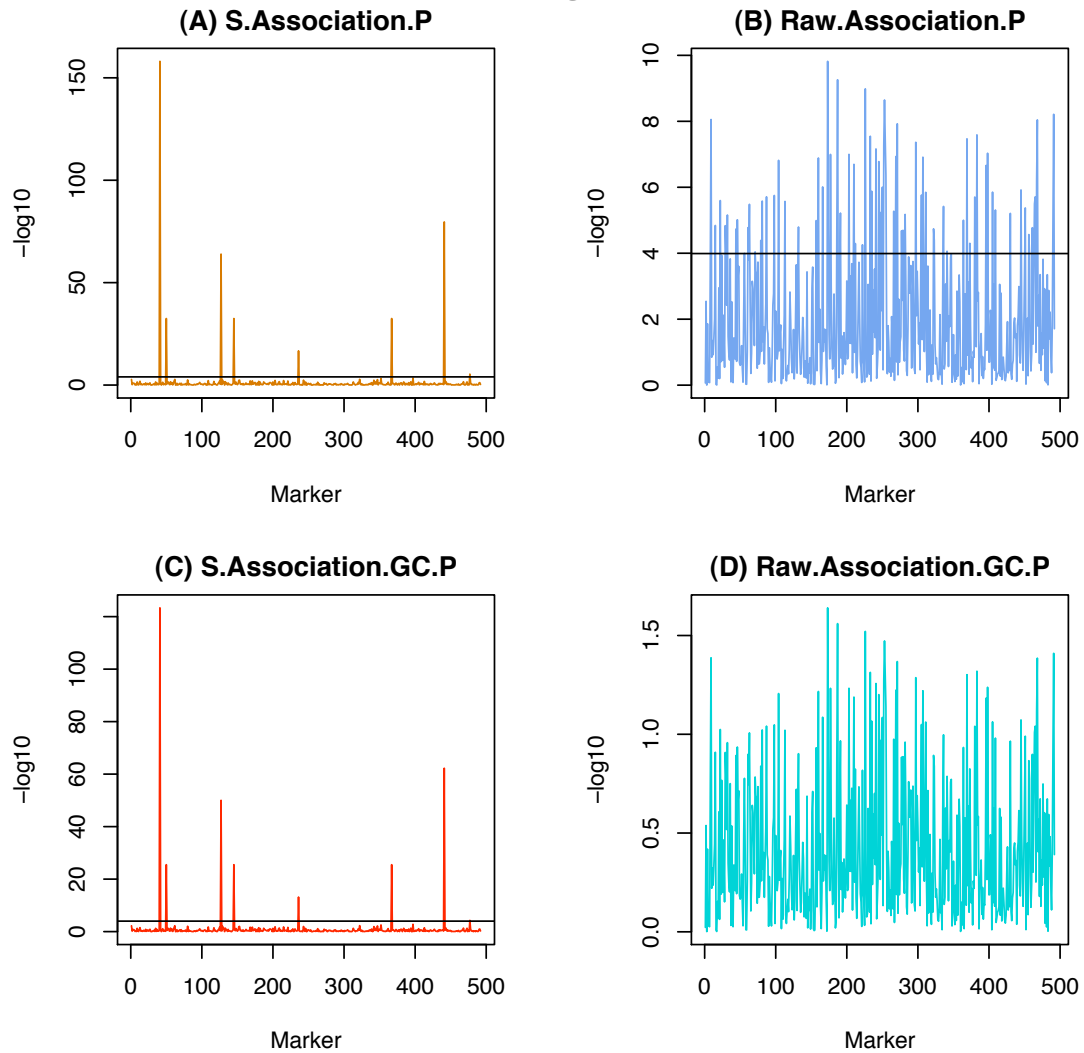


Abbildung 3.27: Assoziationen mit der Blütengröße und der Annahme $K=10$, berechnet in R 2.8.1. Der $-\log_{10}$ transformierte p-Wert ist auf der y-Achse und die zu testenden Marker sind entlang der x-Achse aufgetragen. Das bonferroniadjustierte Signifikanzniveau ist bei 3,99 als horizontale schwarze Linie dargestellt. Nähere Erläuterungen sind dem Text Kapitel 3.2.4.1 zu entnehmen

Tabelle 3.9: **Zusammenfassung: Assoziationsstudie mit R 2.8.1** In dieser Tabelle ist die Anzahl der signifikant assoziierten Marker mit der entsprechenden Eigenschaft über die mögliche Anzahl an Subpopulationen, berechnet mit R 2.8.1, aufgeführt.

Eigenschaft	K=10	K=11	K=12
Herkunft	11	11	19
Blütenfarbe	5	9	17
Blütentyp	11	15	17
Blütengröße	8	9	13
Blütenkonsistenz	7	8	12
Blattform	14	12	21
Blatttyp	10	17	21
Blattzeichnung	7	10	13
Anthocyane im Blatt	5	16	17
Rispenlänge	12	12	23
Duft	12	12	23

3.2.4.2 Assoziationsstudie mit TASSEL

In Tabelle 3.10 ist eine Zusammenfassung der Anzahl an Markern aufgeführt, die eine signifikante Assoziation zur untersuchten Eigenschaft zeigen. Die Signifikanzschwelle liegt hier bei $p=0,05$. In Klammern ist die Anzahl der Marker aufgeführt, die auch bei einem bonferroniadjustierten p -Wert von $p=0,05/492$ eine Assoziation zur untersuchten Eigenschaft zeigen.

Tabelle 3.10: **Zusammenfassung: Assoziationsstudie mit TASSEL 2.1** In dieser Tabelle ist die Anzahl der signifikant assoziierten Marker mit der entsprechenden Eigenschaft über die mögliche Anzahl an Subpopulationen, berechnet mit TASSEL 2.1, aufgeführt. In Klammern ist die Anzahl der Marker aufgeführt, die auch bei einem Bonferroni-adjustierten p-Wert signifikante Assoziationen zur entsprechenden Eigenschaft aufweisen.

Eigenschaft	K=10	K=11	K=12
Herkunft	47 (0)	30 (0)	40 (0)
Blütenfarbe	27 (0)	30 (0)	25 (0)
Blütentyp	33 (0)	31 (1)	32 (0)
Blütengröße	45 (0)	48 (0)	51 (0)
Blütenkonsistenz	60 (2)	42 (1)	67 (1)
Blattform	43 (0)	42 (0)	40 (0)
Blatttyp	36 (0)	35 (0)	42 (0)
Blattzeichnung	37 (0)	34 (0)	40 (0)
Anthocyane im Blatt	31 (0)	34 (0)	36 (0)
Rispenlänge	39 (3)	42 (3)	38 (3)
Duft	51 (6)	51 (6)	53 (6)

3.2.4.3 Vergleich der Ergebnisse der Assoziationsstudie berechnet in R 2.8.1 und TASSEL 2.1

Mit dem Programm TASSEL 2.1 konnten mehr signifikant assoziierte Marker pro Eigenschaft verifiziert werden, als mit R 2.8.1. Jedoch nur, wenn das Signifikanzniveau nicht durch eine Bonferroni-Adjustierung hochgesetzt wurde. Vergleicht man die Ergebnisse aus R 2.8.1 mit den Ergebnissen aus TASSEL 2.1 so ist zu sehen, daß bei einem bonferroniadjustierten p-Wert von 0,05/492, die Anzahl signifikant assoziierter Marker in TASSEL 2.1 wesentlich geringer ist.

Tabelle 3.11 zeigt die Marker, die in beiden Methoden signifikante Assoziationen zu dem jeweiligen Merkmal zeigten.

Tabelle 3.11: Marker, die in den Berechnungen mit R 2.8.1 und TASSEL 2.1 für das jeweilige Merkmal eine signifikante Assoziation zeigten; Marker, die eine Assoziation in mehreren Eigenschaften zeigen sind farblich markiert.

Anthocyangehalt im Blatt	H700AAT_MACAG_17 H700CAT_MAGAG_12 H700CAT_MAGTC_4 H800ATA_MACGT_21	Blattzeichnung	H700AAT_MACAG_5 H700CAA_MACGA_3 H700CAT_MCAAC_4 H800ATA_MACGT_25 H800ATT_MAAGC_3
Blütenfarbe	H700AAT_MACAG_5 H700CAT_MCAAC_2 H700CGA_MACAT_24 H700CGA_MACAT_4 H700CGA_MACGT_24 H800ATT_MAAGC_3	Blattform	H700AAT_MACAG_5 H700CAA_MACGA_25 H700CAT_MCAAC_1 H700CGA_MACGT_4 H800ATT_MAAGC_3
Blüten- konsistenz	H700CAA_MACAC_2 H800ACA_MACAT_2 H800ATA_MAAGC_1	Blatttyp	H700CAA_MACGA_24 H700CAT_MAGAG_2 H700CGA_MACAT_2 H800ATT_MAAGC_18
Blütengröße	H700AAT_MACAG_5 H800ATT_MAAGC_3	Herkunft	H700CAT_MACAG_25 H700CAT_MCAAC_1 H800ACA_MAGTC_18 H800ATA_MAAGC_1 H800ATA_MACGT_21
Blütentyp	H700CAA_MACAC_5 H800ACA_MAGTC_7 H800ATT_MAAGC_3		
Duft	H700CAA_MACAC_31 H700CAA_MACGA_1 H700CAT_MACAT_4 H700CAT_MAGAG_2 H700CGA_MACAT_20 H700CGA_MACAT_4 H800ACA_MACAG_2	Rispenlänge	H700CAA_MACAC_5 H700CAA_MACGA_1 H700CAT_MAGAG_12 H700CAT_MAGAG_14 H700CGA_MACAT_20 H700CGA_MACGT_24 H800ACA_MACAG_2 H800ATA_MAAGC_1 H800ATT_MAAGC_3 H800ATT_MACAT_17

3.3 Verknüpfung der familienbasierten QTL-Analyse und der genomweiten Assoziationsstudie

Die Eigenschaften Blütengröße und Rispenlänge wurden sowohl in der Nachkommenschaft der Kreuzung '21x19', als auch in der Genotypensammlung erfasst. Ein Vergleich der Ergebnisse beider Studien zeigt einige Unterschiede.

In Abbildung 3.28 sind alle Kopplungsgruppen aufgeführt, die einen Einfluss auf die Eigenschaft Blütengröße zeigen. Abbildung 3.29 zeigt die Kopplungsgruppen, die einen Einfluss auf die Eigenschaft Rispenlänge haben. Die Beschriftung der Kopplungsgruppen besteht aus der Nummer der Kopplungsgruppe, äquivalent zur Bezeichnung wie sie in der maternalen wie auch paternalen Karte benutzt wurden (siehe Abbildung 3.7 und 3.6), und dem Buchstaben M bzw V für die Herkunft der Kopplungsgruppe aus der maternalen bzw. paternalen Karte.

Rechts neben der Kopplungsgruppe ist der Graph des Intervallmappings (berechnet mit MapQTL6 und dargestellt mit MapChart) aufgeführt. Die gestrichelte Linie stellt die Chromosomen- bzw. Kopplungsgruppenweite Signifikanzschwelle dar und die durchgezogene Linie zeigt die genomweite Signifikanzschwelle.

Neben der Markerbezeichnung befinden sich an einigen Markern Sternchen(*) in unterschiedlicher Anzahl. Diese Sternchen stehen für die Signifikanz des Einflusses des Markers auf das Merkmal, berechnet mit Kruskal-Wallis in MapQTL 6. Dabei werden von MapQTL 6 folgende Signifikanzniveaus ausgegeben: *: 0.1; **: 0.05; ***: 0.01; ****: 0.005; *****: 0.001; *****: 0.0005; *****: 0.0001. Um die Signifikanz und somit die Qualität des Kruskal-Wallis Tests für die Berechnung des Einflusses des Markers auf das Merkmal zu erhöhen, wurde das Signifikanzniveau um die Anzahl der Chromosomen/Kopplungsgruppen erhöht. Damit liegt der p-Wert bei $p=0,05/19=0,003$. Da die Ausgabe bei MapQTL 6 mit drei Sternen ($p=0.005 \rightarrow$ zu niedrig) oder fünf Sternen ($p=0.001 \rightarrow$ zu signifikant) zu ungenau ist, wurde der K^* -Wert als Teststatistik benutzt. In Tabelle 3.7 sind alle Marker aufgeführt, die diese Signifikanzschwelle für das entsprechende Merkmal überschritten haben. Diese Marker sind in Abbildung 3.28 und 3.29 rot markiert.

Grün markierte Marker sind Marker, die in der Assoziationsstudie einen signifikanten Einfluss auf das entsprechende Merkmal entweder in der Berechnung mit R 2.8.1 oder mit TASSEL 2.1 zeigten. Es sind in Abbildung 3.28 und 3.29 nur Kopplungsgruppen dargestellt, die Marker tragen, welche einen Einfluss in mindestens einer der angewandten Methoden zeigten. In beiden Abbildungen ist zu sehen, dass Marker, die einen Einfluss auf das Merkmal in der genomweiten Assoziationsstudie zeigten, keinen signifikanten Einfluss in der familienbasierten QTL-Studie haben.

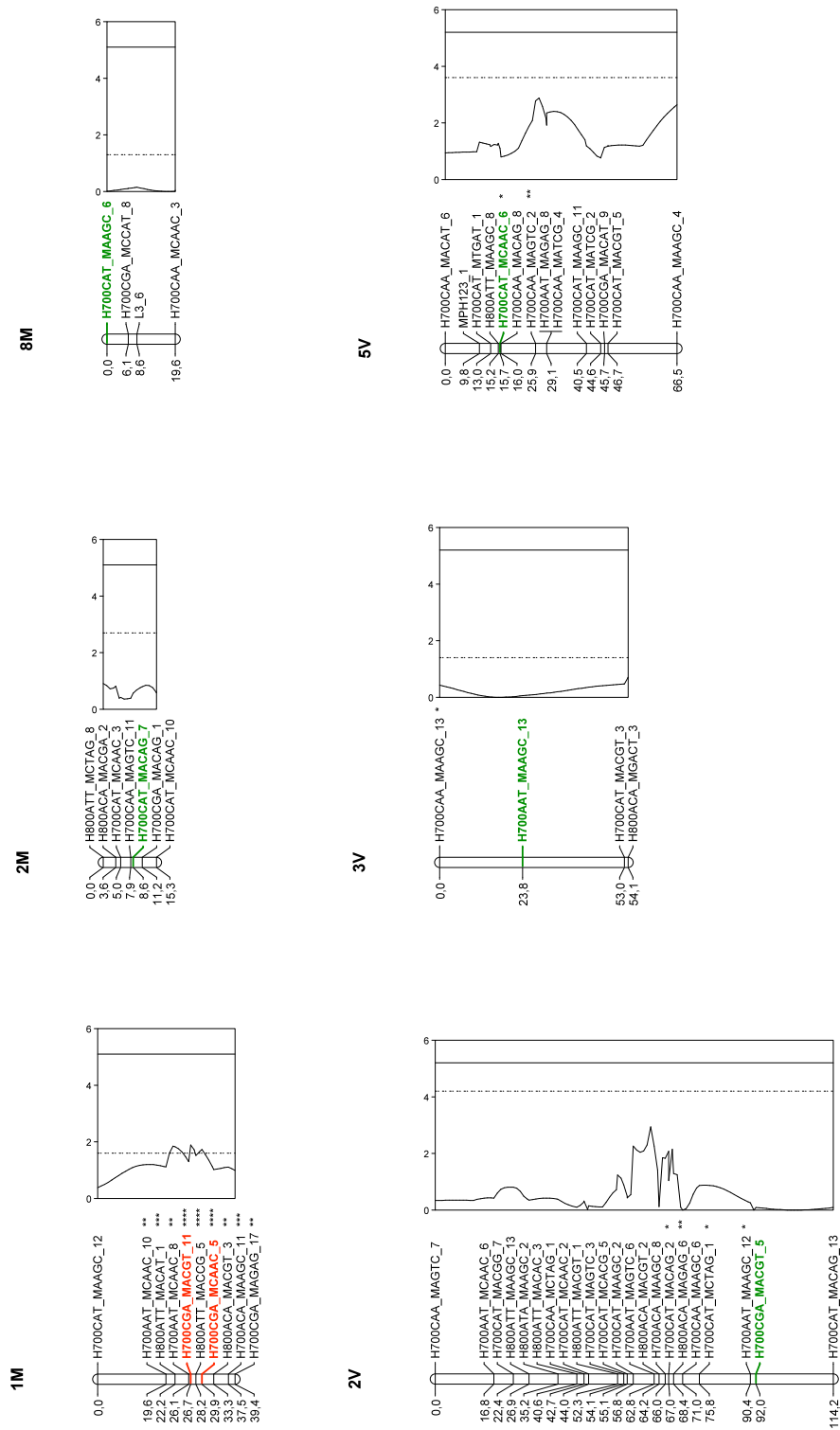


Abbildung 3.28: QTL (rot) und Einzel-Marker-Eigenschafts-Assoziationen (grün) für die Eigenschaft Blütengröße (Eine nähere Beschreibung der Graphen ist dem Text zu entnehmen)

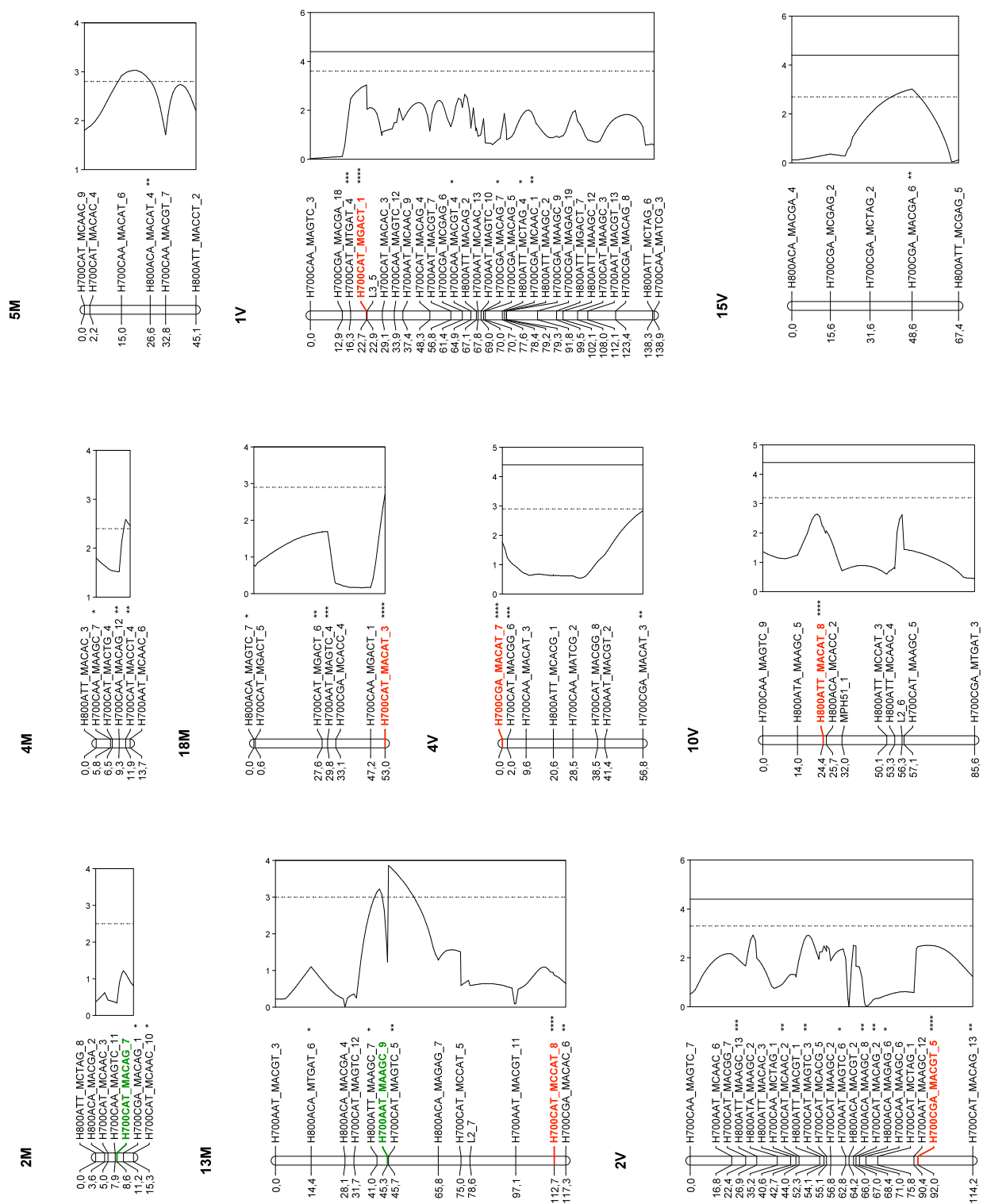


Abbildung 3.29: QTL (rot) und Einzel-Marker-Eigenschafts-Assoziationen (grün) für die Eigenschaft Rispenlänge (Eine nähere Beschreibung der Graphen ist dem Text zu entnehmen)

4 Diskussion

4.1 Kartierung gartenbaulich wichtiger Merkmale

4.1.1 Phänotypische Merkmale in der Nachkommenschaft

In der Nachkommenschaft der Kreuzung '21x19' wurden die Merkmale 'Anthocyangehalt in der Blüte' (Tepalen und Lippe), 'Anthocyangehalt im Blatt', 'Rispenlänge' und 'Blütengröße' aufgenommen. Um eventuelle Zusammenhänge zwischen dem Anthocyangehalt in verschiedenen Pflanzenorganen sowie der Blütengröße und Rispenlänge zu ermitteln, wurde der Korrelationskoeffizient zwischen diesen Merkmalen berechnet.

Wie in Abbildung 3.5 (a) dargestellt, beträgt der Korrelationskoeffizient zwischen dem Anthocyangehalt im Blatt und in der Blüte 0,1785. Da dieser Korrelationskoeffizient relativ niedrig ist, lässt sich daraus schließen, dass es keine Korrelation zwischen dem Anthocyangehalt im Blatt und dem Anthocyangehalt in der Blüte in der vorliegenden Population gibt. In der untersuchten Nachkommenschaft, kann man demzufolge nicht von der Färbung des Blattes auf die Ausprägung der Blütenfarbe schließen.

Der Korrelationskoeffizient zwischen Blütengröße und Rispenlänge beträgt 0,3182 und ist in Abbildung 3.5 (b) dargestellt. Auch dieser Korrelationskoeffizient ist relativ niedrig. Jedoch lässt sich hier eine Tendenz absehen. Große Blüten konnten eher an langen Rispen, aber auch an kurzen Rispen beobachtet werden, während kleinere Blüten nur an kurzen Rispen zu finden waren.

Es muss allerdings beachtet werden, dass die Probenentnahme zu unterschiedlichen Zeitpunkten erfolgte und das unterschiedliche Alter der Pflanzen bei Probennahme. Chalker-Scott (1999) zeigt auf, dass Umwelteinflüsse wie z.B. Lichteinstrahlung, Temperatur- und Wasserstress, sowie das Alter der Pflanzen die Bildung von Anthocyanen im pflanzlichen Gewebe stark beeinflussen. Die untersuchten Pflanzen wurden nicht in einer kontrollierten Umgebung kultiviert, daher waren die Pflanzen Schwankungen der Umwelt unterworfen. Des Weiteren wurden bei einigen Pflanzen Symptome beobachtet, die auf einen Virusbefall schließen lassen. Dixon et al. (1994) zeigten auf, dass auch ein Pathogenbefall einen signifikanten Einfluss auf die Bildung von Anthocyanen sowie auf den Habitus einer Pflanze haben kann.

4.1.2 Molekulare Marker

Für die in der vorliegenden Arbeit durchgeführten Analysen wurden zwei Markersysteme verwendet. Zum einen wurden neue Mikrosatellitenmarker entwickelt und verfügbare Mikrosatellitenmarker genutzt. Zum anderen wurden AFLP-Marker an das Genom von *Phalaenopsis* Hybriden angepasst und für die Analysen verwendet. Die Entwicklung, Anpassung und Nutzung dieser beiden Markersysteme auf das *Phalaenopsis* Genom wird im Folgenden erläutert.

4.1.2.1 Entwicklung von Mikrosatellitenmarkern

In der vorliegenden Arbeit wurden insgesamt 17 Mikrosatelliten Primerpaare für *Phalaenopsis* Hybriden entwickelt. Zu Beginn der Arbeiten im Mai 2006 waren noch keine Mikrosatelliten Primerpaare für *Phalaenopsis* entwickelt worden, bzw. frei zugänglich. In der Zwischenzeit wurden einige wenige Arbeiten zur Entwicklung von Mikrosatelliten für *Phalaenopsis* durchgeführt und veröffentlicht (Lin (2002) seit 2007 öffentlich zugänglich, Rodrigues and Kumar (2008)).

Zur Entwicklung von Mikrosatellitenmarkern wurde eine mit Mikrosatellitenmotiven angereicherte DNA-Bank erstellt. Aus insgesamt 3267 Klonen wurden 441 Mikrosatellitenmotiv-tragende Klone identifiziert. In verschiedenen Anreicherungsrounds wurden Anreicherungs-effizienzen zwischen sieben und 16,9% erreicht (siehe Tabelle 3.2). Im Vergleich dazu wurden beispielsweise aus einer DNA-Bank für *Actinidia chinensis* nur 0,1 bis 1,2% Klone mit Mikrosatellitenmotiv identifiziert (Weising et al. (1996)). Demzufolge konnte die Ausbeute von Mikrosatellitenmotiv-tragenden Fragmenten durch die Anreicherung wesentlich erhöht werden.

Bei der Anreicherung in *Eucalyptus* sp. wurde eine Anreicherungs-effizienz von 10 bis 40% erreicht (Brondani et al. (1998)), in *Gossypium hirsutum* lag die Anreicherungs-effizienz zwischen 13,3 und 24,6% (Kumpatla et al. (2004)) und in *Pelargonium* sp. konnte eine Effizienz von 5 bis 39% erreicht werden (Becher et al. (2000)). Damit liegt die Anreicherungs-effizienz in der vorliegenden Arbeit eher im unteren Bereich. Allerdings war in der vorliegenden Arbeit der Anteil Klone, aus denen effektiv Mikrosatellitenmarker konstruiert werden konnten, höher als in den zuvor genannten Arbeiten. Es stellte sich heraus, dass der Arbeits- und Zeitaufwand von der Identifizierung positiver Klone bis zur Entwicklung funktionaler Mikrosatellitenmarker sehr groß ist (Squirrell et al. (2003)). Daher ist die Qualität der angereicherten DNA-Fragmente wichtiger als eine hohe Klonzahl, die unter Umständen einen großen Anteil an Duplikaten oder Chimären enthalten kann.

Aus einer großen Anzahl von Klonen konnten nur wenige funktionierende Mikrosatelliten Marker entwickelt werden. Squirrell et al. (2003) zeigen Gründe auf, die zu

einer Abnahme der Anzahl von Mikrosatellitenmotiven bzw. -markern führt. Zunächst sind in angereicherte DNA Banken nicht nur Klone, die ein Fragment mit Mikrosatellitenmotiv enthalten, vorhanden. Um die relevanten Klone zu identifizieren, sind verschiedene Strategien möglich. In der vorliegenden Arbeit wurden Klone, die ein Fragment mit Mikrosatellitenmotiv tragen, durch Hybridisierung mit komplementären Sonden identifiziert. Da nur Klone, die ein starkes Signal gaben, für die weitere Analyse verwendet wurden, war die Selektion relativ stringent und es besteht somit die Möglichkeit einige positive Klone frühzeitig ausgeschlossen zu haben. Eine weitere Reduktion der Klone geschah durch die Größenfraktionierung der Inserts. Da nur Klone mit einer Insertgröße von mindestens 500bp verwendet wurden, wurden auch hier wieder potentielle Mikrosatellitensequenzen verworfen.

Squirrell et al. (2003) zeigen weiterhin auf, dass Protokolle mit einer hohen Anreicherungs-effizienz eine hohe Anzahl an chimären Klonen oder Duplikate aufweisen. Obwohl in der vorliegenden Arbeit die Anreicherungs-effizienz relativ niedrig ist, konnten aufgrund der hohen Selektionsstringenz die anschließenden Reduktionsfaktoren verringert werden. So zeigten 85,6% der sequenzierten Klone Mikrosatellitenmotive und es zeigte sich mit nur einem Duplikat eine sehr niedrige Redundanz. Aufgrund der Größenfraktionierung konnten aus nahezu allen Sequenzen (92,7%) Primerpaare entwickelt werden. Es gab nur wenige Sequenzen, deren flankierende Regionen zu klein oder ungeeignet für das Primerdesign waren.

In der letzten Phase vom Primerdesign zum funktionalen Mikrosatellitenmarker kommt es nochmals zu einer starken Reduktion. Im Durchschnitt mehrerer von Squirrell et al. (2003) zusammengefasster Arbeiten zeigten 19,5% der entwickelten Primerpaare keine Amplifikate und durchschnittlich 24% ein komplexes Muster mit mehreren Banden, das nicht interpretierbar war. Wie von den Autoren aufgeführt, können Gründe für das Ausfallen der Amplifikation, chimäre Klone oder Fehler bzw. Mutationen in den zugrunde liegenden Sequenzen sein. Die Amplifikation mehrerer nicht interpretierbarer Banden kann auf eine große Menge an repetitiver DNA im Genom zurück zu führen sein. Eine weitere Reduktion wird durch die große Anzahl monomorpher Mikrosatelliten bedingt. Zum Test auf Polymorphie der Mikrosatellitenfragmente wird eine Auswahl möglichst unterschiedlicher Genotypen verwendet. In der vorliegenden Arbeit wurde ein Mikrosatellitenmarker als polymorph bezeichnet, wenn mindestens ein Unterschied in einer Auswahl aus zwölf *Phalaenopsis* Hybriden und zwei Wildarten auftrat. Dadurch wurde die Anzahl möglicher Mikrosatellitenmarker nochmals um 50% reduziert.

In der vorliegenden Arbeit wurden die DNA Banken nach den Motiven (AG)_n und (GT)_n angereichert. Diese Motive wurden für die Anreicherung gewählt, da sie mit am häufigsten im Genom vorkommen (Weising et al. (2005)). Die höhere Anreicherungs-

effizienz des Motivs $(AG)_n$ lässt sich dadurch erklären, daß das Motiv $(AG)_n$ häufiger als das Motiv $(GT)_n$ im pflanzlichen Genom vorkommt (Weising et al. (2005)).

Mikrosatelliten, die Polymorphismen zeigten, wurden schließlich für die Analyse des *Phalaenopsis* Genoms verwendet. Als Maß für die Qualität eines Mikrosatellitenmarkers wird der 'Polymorphic Information Content' (PIC) verwendet. Obwohl sich der PIC und Heterozygotiegrad nur gering unterscheiden (PIC ist immer etwas kleiner als der Heterozygotiegrad) werden sie häufig verwechselt und der Heterozygotiegrad als PIC angegeben. Shete et al. (2000) definieren die Heterozygotie als „the propability that a random individual chosen from the population is heterozygous at a locus“, dagegen definieren sie den PIC-Wert als „the probability that a given marker genotype of an offspring of an affected parent will allow deduction of the parental genotype at the Marker locus“. Die Unterschiede in den Berechnungen sind in Formel 2.1 und 2.2 aufgeführt. Der PIC Wert und der Heterozygotiegrad für die in dieser Arbeit verwendeten Mikrosatellitenmarker ist in Tabelle 3.4 aufgeführt. Der PIC-Wert liegt zwischen 0,35 und 0,85 mit einem Mittelwert von 0,7. Der durchschnittliche Heterozygotiegrad in der vorliegenden *Phalaenopsis* Population beträgt 0,73. Laut Miller and Yuan (1997) sind PIC-Werte über 0,7 üblich. In verschiedenen Studien wurde jedoch ein relativ weiter Bereich von PIC-Werten gefunden. In einer Studie von Smith et al. (1997) an *Zea* wurden Mikrosatellitenmarker mit einem PIC von 0,06 bis 0,91 gefunden. In *Oryza* hingegen reichte der PIC von 0,744 bis 0,957 (Blair et al. (2002)). In der Arbeit von Lin (2002) zur Isolierung von Mikrosatelliten in *Phalaenopsis* wurde ein PIC von 0,19 bis 0,7 festgestellt. Hierbei wird deutlich, dass der PIC-Wert nicht nur vom Polymorphiegrad des Markers, sondern auch von der Größe der Population abhängt. So hat der Mikrosatellitenmarker 'L3' in der Arbeit von Lin (2002) nur einen Wert von 0,43 mit einer Testpopulation von 14 bis 15 *Phalaenopsis* sp. In der vorliegenden Arbeit mit einer Populationsgröße von 144 verschiedenen Genotypen lag der PIC-Wert hingegen bei 0,84.

Eine Anreicherung von DNA-Fragmenten zur Isolierung von Mikrosatellitenmotiven wurde auch bereits bei einem anderen Vertreter der Familie *Orchidaceae* durchgeführt. In der Arbeit von Yue et al. (2006) wurden vierzehn polymorphe Mikrosatellitenmarker mit einem durchschnittlichen Heterozygotiegrad von 0,7 und sechs bis 33 Allele innerhalb von 42 *Dendrobium* Hybriden entwickelt. In der vorliegenden Arbeit amplifizierten die verwendeten Mikrosatellitenmarker in einer *Phalaenopsis*-Population mit 144 verschiedenen Genotypen zwischen zwei und 25 Allele und zeigten einen Heterozygotiegrad von 0,73. Des Weiteren konnten mit Hilfe von sechs Mikrosatelliten, in einer Neighbour Joining Analyse, 136 von 144 unterschiedlichen Genotypen eindeutig voneinander unterschieden werden (siehe Anhang Abbildung 5.1). Die relativ hohen PIC-Werte, sowie die Möglichkeit eine hohe Anzahl von Genotypen mit

wenigen Mikrosatelliten eindeutig unterscheiden zu können, zeigt die gute bis sehr gute Qualität der entwickelten Mikrosatellitenmarker. In der praktischen *Phalaenopsis* Züchtung können diese Marker zur Unterscheidung von Genotypen angewandt werden. Des Weiteren ist eine Anwendung der Marker zur Erstellung einer genetischen Karte sowie für genomweite Assoziationsstudien möglich.

In der Anwendung der Mikrosatellitenmarker ist aufgefallen, dass die PCR Reaktionen mit DNAs der Wildarten häufiger negativ ausfielen als die PCR Reaktionen mit DNAs der Hybriden. Des Weiteren konnte ein Großteil der von Lin (2002) und Rodrigues and Kumar (2008) veröffentlichten Marker aufgrund fehlender Amplifikate nicht genutzt werden. Diese Mikrosatelliten wurden allerdings in Wildarten entwickelt. Eine Erklärung hierfür kann sein, dass sich die flankierenden Sequenzen der Wildarten so stark von denen der Hybriden unterscheiden, dass eine Anlagerung der Primer nicht möglich ist. Da die Wildarten auch im Dendrogramm eine relativ große Distanz zu den Hybriden zeigen (Abbildung 3.15), scheinen die flankierenden Sequenzen nicht konserviert genug zu sein, um bei Wildarten und Hybriden gleichermaßen gut zu funktionieren.

4.1.2.2 Anpassung des AFLP-Markersystems

Ein entscheidender Schritt zur Anpassung der AFLP-Technik an das *Phalaenopsis*-Genom war die Erweiterung der selektiven Basen auf sieben, statt, wie von Vos et al. (1995) beschriebenen, sechs Basen in der selektiven Endreaktion. Durch die Erweiterung der Anzahl selektiver Basen konnte die Selektivität der Endreaktion erhöht und die Anzahl der amplifizierten Fragmente verringert werden. Dies erhöhte die Qualität der Gelbilder und eine eindeutige Auswertung war möglich. Die Erhöhung der Anzahl der selektiven Basen ist eine seit Jahren besonders bei Arten mit großen Genomen erfolgreich angewandte Methode (Remington et al. (1999), Han et al. (1999), Mueller and Wolfenbarger (1999)).

4.1.3 Erstellung der genetischen Karte

Die in dieser Arbeit erstellte genetische Karte ist die erste Karte für eine Gattung aus der Familie *Orchidaceae* und demzufolge auch die erste genetische Karte für *Phalaenopsis* Hybriden. Für die Erstellung dieser Karte war nur eine polyploide Kartierungspopulation verfügbar. Aufgrund cytologischer Untersuchungen liegt die Vermutung nahe, daß es sich dabei um eine tetraploide Population handelt. Da jedoch der Ploidiegrad nicht genau bestimmt werden konnte, und auch die Segregationsmuster der Mikrosatellitenanalyse nicht mit dem von Tetraploiden übereinstimmen, wurde eine Kartierungsmethode gewählt, die unabhängig vom Ploidiegrad der Kartierungspopu-

lation ist. Die Verwendung von Single Dose Markern (SDM) wurde bereits erfolgreich bei beispielsweise *Salix* (Barcaccia et al. (2003)), *Medicago* (Julier et al. (2003)) und *Paspalum* (Stein et al. (2007)) zur Erstellung einer genetischen Karte in einer polyploiden Nachkommenschaft genutzt. Selbst in octoploiden Organismen wie *Fragaria* ist die Erstellung einer genetischen Karte mit SDM möglich (Lerceteau-Köhler et al. (2003)).

In Abbildung 3.6 und 3.7 ist jeweils die paternale und maternale genetische Karte der Nachkommenschaft der Kreuzung '21x19' zu sehen. Beide Karten decken nur jeweils ca. 30% des Genoms ab. Diese geringe Abdeckung ist durch die relativ kleine Anzahl der kartierten Marker gegenüber der durchschnittlichen Größe des *Phalaenopsis* Genoms erklärbar. Die relativ geringe Anzahl der für die Kartierung detektierten Marker wurde nochmals reduziert, indem ausschließlich SDM verwendet wurden. Wu et al. (1992) schlagen vor, dass 75 Pflanzen für eine Identifizierung von SDM mit einem 98%igen Konfidenzintervall in tetraploiden Pflanzen ausreichend sind. In der vorliegenden Arbeit basiert die Kartierung auf 103 Pflanzen. Demnach ist von einer ausreichend hohen Absicherung der identifizierten und für die Kartierung verwendeten SDM auszugehen. Um eine bessere Abdeckung des Genoms zu erreichen, ist es notwendig die Anzahl der SDM entscheidend zu erhöhen. Die für die Kartierung verwendete Population zeigte jedoch kaum Polymorphismen und folglich waren relativ viele AFLP Primerkombinationen (89) notwendig, um die in dieser Arbeit verwendete Anzahl an polymorphen Markern zu erhalten. Um jedoch eine Kartierungspopulation zu generieren, bei der die Wahrscheinlichkeit polymorphe SDM zu detektieren erhöht ist, sollte ein Elter für eine hohe SDM Frequenz eine hohe Heterozygotie (z.B. out-cross hybrid) und der zweite Elter eine geringe Heterozygotie (z.B. Doppelhaploide oder Nachkomme aus Inzucht) aufweisen (Wu et al. (1992)).

Die geringe Anzahl an Ankermarkern zeigt ebenfalls die zu geringe Abdeckung des kartierten Genoms. Es konnte keine Verknüpfung zwischen den einzelnen Kopplungsgruppen der paternalen und maternalen Karte durchgeführt werden, da zu wenig biparental vererbte Marker vorhanden waren. Aufgrund der acht gefundenen Ankermarker (siehe Tab 3.6) lässt sich vermuten, dass folgende Kopplungsgruppen (LG) zusammen gehören (jeweils erst die paternale und dann die mögliche homologe maternale Kopplungsgruppe): LG2 zu LG4, LG5 zu LG5, LG6 zu LG14 und LG15 zu LG13. Auf der paternalen Kopplungsgruppe LG14 befinden sich drei Ankermarker, die in der maternalen Karte auf den Kopplungsgruppen LG9 und LG5 zu finden sind. Dass einige Ankermarker mal einer Kopplungsgruppe und dann wieder verschiedenen anderen Kopplungsgruppen zugeordnet werden, zeigt die schwache Markerabsättigung der genetischen Karte. Die unterschiedliche Zuordnung der Ankermarker ist ein Anzeichen dafür, dass einzelne Kopplungsgruppen durchaus auf ein und demselben

Chromosom liegen können. Demnach können in der vorliegenden Karte getrennte Kopplungsgruppen bei einer höheren Markerabsättigung zusammen fallen und neue, noch nicht erfasste Kopplungsgruppen gebildet werden. In der vorliegenden genetischen Karte sind bisher nur anonyme Marker kartiert worden. Für eine Verlinkung der Kopplungsgruppen der paternalen und maternalen Karte ist neben einer Erhöhung der Markerdichte auch die Kartierung von spezifischen / nicht anonymen Markern (z.B. EST-Marker, Marker für Strukturgene,...) hilfreich. Der Vorteil von diesen Markern ist, dass durch sie Kandidatengene identifiziert werden können, da sie den direkten Zugriff auf einzelne Gene, die eine bestimmte Funktion in der Pflanze erfüllen, ermöglichen. Auch die Implementierung anderer Markertypen wie z.B. SNP's (Single Nucleotide Polymorphism) oder SAMPL (Selective Amplification of Microsatellite Polymorphic Loci) Marker kann für eine Verdichtung der Karte genutzt werden.

Die Erstellung einer hochauflösenden Karte mit der vorhandenen Population ist schwierig. Wie bereits erwähnt, wurden in der vorliegenden Arbeit 103 Nachkommen für die Kartierung verwendet. Weising et al. (2005) geben 100 Nachkommen als untere Grenze für die Detektion von QTL an und eine Populationsgröße von 500 oder sogar 1000 Nachkommen für die Erstellung einer hochauflösenden Karte. Folglich kann die erstellte Karte mit der verwendeten Kartierungspopulation durch zusätzliche Marker noch verdichtet werden, jedoch ist für eine hochauflösende Karte, wie sie zum Beispiel für *Triticum* (Quarrie et al. (2005)), *Papaya* (Chen et al. (2007)) oder *Hordeum* existiert, eine wesentlich höhere Anzahl von Nachkommen notwendig. Des Weiteren zeigen Hackett et al. (1998) auf, dass es aufgrund der relativ geringen Populationsgröße zu einer Unterschätzung der Gesamtlänge der Karte in den Berechnungen mit JoinMap kommen kann. Auch die Reihenfolge der kartierten Marker ist kritisch zu betrachten. Ferreira et al. (2006) zeigen auf, dass die Anzahl der Genotypen einer Kartierungspopulation einen starken Einfluss haben kann. So kann die Reihenfolge der Marker beeinflusst sein und aus einer Kopplungsgruppe fälschlicherweise mehrere Kopplungsgruppen entstehen.

All die genannten Faktoren zeigen auf, daß die erstellte Karte als erste genetische Karte für *Phalaenopsis* noch stark verbesserungswürdig ist. Für eine Verbesserung der Kartendichte und -struktur sollte jedoch eine Population mit deutlich segregierendem Phäno- und Genotyp, sowie einer höheren Anzahl von Nachkommen verwendet werden. Allerdings gibt die in dieser Arbeit erstellte Karte bereits einen guten Überblick über die genomische Struktur von *Phalaenopsis* und kann als Grundlage für weitere Analysen und zur Kartierung von gartenbaulich wichtigen Merkmalen dienen. Auch konnte die erstellte Karte in der vorliegenden Arbeit bereits zur Detektion von QTL genutzt werden und somit Hinweise auf bestimmte Regionen, welche für ein Merkmal verantwortlich sind, gefunden werden.

4.1.4 QTL-Analyse

Die QTL-Analyse ist ein leistungsstarker Ansatz, um gartenbaulich interessante Gene zu identifizieren. Für *Phalaenopsis* Hybriden spielt vor allem der Zierwert der Blüte eine große Rolle in der Züchtung neuer Sorten. Aus diesem Grund wurde in der vorliegenden Arbeit die Identifizierung von Genen, die den Zierwert der Blüte aber auch der Gesamtpflanze beeinflussen, angestrebt. Es wurden Merkmale der Blüte wie Anthocyangehalt in den Tepalen und in der Lippe und die Blütengröße, sowie Merkmale der Gesamtpflanze wie Anthocyangehalt in den Blättern und Rispenlänge untersucht. Als Grundlage für die Detektion von QTL wurde die in der aktuellen Studie erstellte genetische Karte (siehe Abbildung 3.6 und 3.7) der Kreuzung '21x19' verwendet. Die Merkmalsausprägungen wurden in der Nachkommenschaft selbiger Kreuzung aufgenommen. Die Verrechnung der Daten erfolgte in MapQTL 6 und die Darstellung der QTL mit MapChart.

Wie in Abschnitt 3.1.4 dargestellt, konnten 13 QTL für die genannten Merkmale in der untersuchten Population detektiert werden. Die Signifikanzschwelle der Teststatistik wurde durch eine Semibonferroniadjustierung aufgrund der Chromosomenzahl, und nicht durch eine Bonferroniadjustierung aufgrund der Gesamtzahl an untersuchten Markern, festgelegt.

Laut Cavanagh et al. (2008) werden vier Voraussetzungen für die Detektion von QTL benötigt. Diese vier Elemente werden im Folgenden beschrieben und im Zusammenhang mit der vorliegenden Studie diskutiert.

Zunächst wird eine Population benötigt, welche sich genetisch für das zu untersuchende Merkmal unterscheidet. Die Untersuchungen in der vorliegenden Arbeit wurden an einer Nachkommenschaft durchgeführt, deren Eltern lediglich in der Blütenfarbe deutliche Unterschiede zeigten. In allen anderen Merkmalen unterscheidet sich die Population nur geringfügig. Trotz der geringen Unterschiede in den Eltern konnten einige, in der Nachkommenschaft segregierende Merkmale gefunden werden. Eine geeignete Nachkommenschaft für eine QTL-Analyse sollte eine hinreichende Variation des zu untersuchenden Merkmals auf genetischem wie auch phänotypischem Level zeigen (Liu (1998)). Es konnten in der Nachkommenschaft zwar Unterschiede in den phänotypischen Ausprägungen einiger Eigenschaften gefunden werden, jedoch waren diese Unterschiede eher gering. Für eine gezielte Untersuchung und Verifizierung der hier gefundenen Assoziationen sollte eine gezielte Kreuzung durchgeführt werden. Des Weiteren wurden relativ viele AFLP Primerkombinationen für die Detektion von Polymorphismen in der Nachkommenschaft benötigt. Dies spricht für eine geringe Variation der Eltern auf DNA Ebene. Um jedoch möglichst viele Polymorphismen in einer Nachkommenschaft zu detektieren, sollten Eltern gewählt werden, die sich

genetisch möglichst weit voneinander unterscheiden. Die Eltern der hier untersuchten Nachkommenschaft unterscheiden sich mit einem Distanzwert von 0,07. Diese relativ geringe genetische Distanz zwischen den Eltern erklärt die hohe Anzahl an Primerkombinationen zur Detektion von Polymorphismen.

Als zweiter Punkt ist die Wahl geeigneter Markersysteme zur Genotypisierung der Population zu nennen. In der vorliegenden Arbeit wurden AFLP und dominant ausgewertete Mikrosatellitenmarker verwendet. AFLP Marker bieten den Vorteil, dass keine Sequenzinformationen der zu untersuchenden Population notwendig sind, da die Fragmente aufgrund bestimmter Schnittstellen im Genom entstehen. Des Weiteren können viele Marker mit einer geringen Anzahl von Primerkombinationen detektiert werden. Mikrosatellitenmarker hingegen müssen für jede Gattung neu entwickelt werden (falls nicht durch andere Gattungen der Familie übertragbar). Allerdings ist der Informationsgehalt eines Mikrosatellitenmarker höher als der eines AFLP Markers, da die Möglichkeit besteht, diesen Marker codominant auszuwerten. Auch wenn die codominante Auswertung der Mikrosatellitenmarker aufgrund der unklaren Ploidiestufe in der vorliegenden Arbeit nicht möglich war, lieferten sie dennoch genug polymorphe Marker. Sie zeigten einen hohen Variationsgrad zwischen den Genotypen. Dies wird bedingt durch die sequenzbasierte und damit locusspezifische Amplifikation von Mikrosatellitenmarkern. Obwohl das AFLP wie auch das Mikrosatelliten Markersystem gut für die Detektion von QTL geeignet ist, kann die Detektion durch eine Kombination mit weiteren sequenzbasierten Markersystemen wie z.B. EST- oder SNP-Markern verbessert werden.

Die Nutzung von reproduzierbaren quantitativen Methoden zur Phänotypisierung sind das dritte Element zur Detektion von QTL. Wie Gene bzw. deren Expression, werden auch QTL und damit die Ausprägung phänotypischer Merkmale stark von der Umwelt beeinflusst (Asins (2002)). Die Aufnahme der quantitativen Daten erfolgte in möglichst vielen Wiederholungen, damit zumindest Messfehler weitestgehend reduziert werden konnten. Allerdings konnten die Daten nur an jeweils einer Pflanze pro Genotyp aufgenommen werden. Des Weiteren wurden auch die Umweltbedingungen nur teilweise kontrolliert. Neben Schwankungen in den Temperaturen kam es außerdem zu einem Virusbefall der Pflanzen. Der Pathogenbefall kann nicht nur einen Einfluss auf den Anthocyangehalt (Dixon et al. (1994)) sondern auch auf den gesamten Habitus der Pflanze haben. Um Messfehler aufgrund des Umwelteinflusses zu verringern, sollten die Daten von jeweils mehreren Klonen pro Genotyp aus mehreren kontrollierten Umwelten aufgenommen werden. Aufgrund der langen reproduktiven Phase von *Phalaenopsis* sollte die Verklonung bereits in der *in vitro* Kultur erfolgen.

Als viertes Element ist die Auswahl der experimentellen und statistischen Methode zur Detektion und Lokalisierung von QTL von entscheidender Bedeutung. In der

vorliegenden Studie konnte keine zufriedenstellende experimentelle Methode in der Wahl der Population oder Aufnahme der phänotypischen Daten erreicht werden, da nur eine Kartierungspopulation vorhanden war und eine gezielte Kreuzung aufgrund der langen reproduktiven Phase von *Phalaenopsis* Hybriden nicht durchgeführt werden konnte. Daher wurde die Stringenz in der statistischen Auswertung der Daten möglichst hoch gesetzt. Als QTL wird ein Locus nur dann bezeichnet, wenn er im Intervallmapping und im Kruskal-Wallis-Test einen signifikanten Einfluss auf das zu untersuchende Merkmal zeigt. Dabei wurde im Kruskal-Wallis-Test eine Semibonferroni-Adjustierung auf Grundlage der Chromosomengrundzahl durchgeführt. In der vorliegenden Studie wurden nur drei Loci gefunden, die beide Bedingungen erfüllen.

Für das Merkmal Blütengröße zeigten auf Kopplungsgruppe 1 der maternalen Karte die Marker H700CGA_MACGT_11 und H700CGA_MCAAC_5 eine signifikante Assoziation im Kruskal-Wallis-Test. Auch im Intervallmapping konnten die umliegenden Regionen die kopplungsgruppenweite Signifikanzschwelle überschreiten. Obwohl diese Schwelle nur knapp überschritten werden konnte, scheint diese Region einen Einfluss auf das Merkmal Blütengröße zu haben, da auch die umliegenden Marker einen Einfluss zeigen.

Der Marker H700CAA_MATCG_8 der Kopplungsgruppe 19 der paternalen Karte zeigt im Kruskal-Wallis-Test einen hoch signifikanten Einfluss auf das Merkmal Anthocyangehalt in der Lippe. Die umliegende Region zeigt im Intervallmapping durch Überschreiten der genomweiten Signifikanzschwelle ebenfalls einen starken Einfluss. Leider zeigt das Intervallmapping genau an der Position des Markers einen Abfall unterhalb der genom- und kopplungsgruppenweiten Signifikanzschwelle. Dies deutet darauf hin, dass die Anordnung der Marker auf dieser Kopplungsgruppe nicht korrekt ist.

Für das Merkmal Anthocyangehalt in der Blüte zeigt der Marker H800ATT_MAAGC_3 einen starken Einfluss im Kruskal-Wallis-Test und auch das Intervallmapping überschreitet hier die kopplungsgruppenweite Signifikanzschwelle. Damit scheint diese Region einen Einfluss auf das Merkmal zu haben. Dieser Marker ist jedoch am Ende einer Kopplungsgruppe lokalisiert, die nur fünf Marker umfaßt. QTL Regionen am Ende einer Kopplungsgruppe deuten auf in der Kartierung fehlende genetische Marker hin. Offensichtlich ist die Markerabsättigung zu gering. Aus diesem Grund kann hier lediglich davon ausgegangen werden, dass diese Region einen Einfluss zu haben scheint, jedoch ist die genaue Lokalisation im *Phalaenopsis* Genom aufgrund der fehlenden Marker noch nicht möglich.

Im Abschnitt 3.1.4 sind weitere Kopplungsgruppen dargestellt, die entweder einen Einfluss im Kruskal-Wallis-Test oder im Intervallmapping auf das untersuchte Merkmal zeigen. Da diese Regionen nur in einem der Tests einen Einfluss zeigen, können

die entsprechenden Regionen lediglich Hinweise auf QTL sein.

Zusammenfassend lässt sich sagen, daß es möglich war QTL bzw. Hinweise auf QTL in der untersuchten Nachkommenschaft zu detektieren. Um diese QTL jedoch zu verifizieren, sollten die Untersuchungen an einer neuen, geeigneteren Nachkommenschaft wiederholt werden. Aufgrund der langen reproduktiven Phase von *Phalaenopsis* sollte die Auswahl der Eltern mit Sorgfalt getroffen und alle oben genannten Punkte in der Herstellung der Population beachtet werden. Werden diese Voraussetzungen erfüllt, so ist es möglich eine genauere Lokalisierung von Genen, die für die Ausprägung eines Merkmals verantwortlich sind, vorzunehmen. Erst dann wird es möglich sein durch markergestützte Selektion die Züchtung von *Phalaenopsis* Hybriden bezüglich der untersuchten Merkmale zu verbessern.

4.2 Assoziationsstudie

Wie viele Methoden in der Pflanzengenetik so stammt auch der Ansatz der genomweiten Assoziationsstudie aus der Humangenetik. Die ersten Analysen erfolgten beim Menschen auf der Suche nach genetischen Markern für Alzheimer (Corder et al. (1994)) und Schizophrenie (Moises et al. (1995)). Wenig später wurde diese Methode für Pflanzen übernommen. Eine Auflistung früher Studien ist in Gupta et al. (2005) zu finden. Inzwischen sind eine Reihe von Assoziationsstudien in verschiedenen Gattungen erfolgreich durchgeführt worden: *Triticum* sp. (z.B. Breseghello (2005)), *Zea mays* (z.B. Remington et al. (2001)), *Solanum tuberosum* (z.B. Gebhardt et al. (2004), D'hoop et al. (2008), Simko (2004)), *Beta vulgaris* (z.B. Stich et al. (2008), Stich et al. (2008)), *Hordeum vulgare* (z.B. Kraakman et al. (2004), Waugh et al. (2009), Cockram et al. (2008)) sowie *Avena sativa* (Achleitner et al. (2008)), *Lolium perenne* (Sköt et al. (2005)) und *Beta vulgaris* ssp. *maritima* (Hansen et al. (2001)). Während in den meisten Studien SNP oder Mikrosatellitenmarker für die Analyse verwendet wurden, wurden in den drei zuletzt genannten Studien, wie in der vorliegenden Arbeit, AFLP Marker verwendet. Eine weitere Auflistung von Assoziationsstudien ist in Zhu et al. (2008) zu finden.

In der vorliegenden Arbeit sollte, als Ergänzung zur familienbasierten QTL-Kartierung, eine Assoziationsstudie zur Detektion gartenbaulich wichtiger Merkmale an einer Sammlung von heute auf dem Markt befindlichen *Phalaenopsis* Hybriden durchgeführt werden. Die entscheidenden Vorteile dieser Methode sind, dass die zu untersuchenden Genotypen nicht miteinander verwandt sein müssen, aufgrund der Ausnutzung mehrerer Rekombinationsereignisse eine höhere Auflösung und damit feinere Kartierung möglich ist und Allele mit nur kleinen Effekten leichter identifiziert werden

können (Oraguzie et al. (2006), Nordborg and Weigel (2008)). Wichtige Voraussetzungen zur Durchführung von Assoziationsstudien ist die Kenntnis über die Struktur der zu untersuchenden Population, die Kenntnis über das Kopplungsungleichgewicht in dieser Population sowie die Aufnahme der entsprechenden phänotypischen Daten. Bevor in Kapitel 4.2.6 die Ermittlung der Marker-Eigenschaft-Assoziationen diskutiert wird, werden in den folgenden Kapiteln zunächst die Populationsstruktur sowie das Kopplungsungleichgewicht der untersuchten *Phalaenopsis* Hybriden erläutert.

4.2.1 Phänotypische Daten in der Genotypensammlung

Das Boniturschema der Genotypensammlung ist in Tabelle 2.1 aufgeführt.

Die Aufnahme der Daten erfolgte an unterschiedlichen Pflanzen eines Genotyps und wurden von der Firma Bock Bio Science GmbH freundlicherweise zur Verfügung gestellt. Die Pflanzen wurden in einem Gewächshaus kultiviert, in dem die Bedingungen nur teilweise kontrolliert werden konnten. Dementsprechend kann der Umweltfaktor nicht in die Berechnungen mit einbezogen, muss allerdings bei der Interpretation der Ergebnisse mit beachtet werden. Des Weiteren war die Aufnahme der Daten relativ ungenau. So wurden beispielsweise Blütenkonsistenz, Anthocyane im Blatt und der Blütenduft nur durch Kategorien wie 'fest / weniger fest' oder 'ja / nein' erfasst. Diese Art der Datenaufnahme ist sehr ungenau und unter Umständen sehr subjektiv.

Eine bessere Phänotypisierung ist für die Verifizierung der in dieser Studie gewonnenen Ergebnisse von dringender Notwendigkeit. So sollten zum einen mehrere Klone eines Genotyps in unterschiedlichen Umwelten kultiviert werden und über mehrere Vegetationsperioden phänotypisiert werden. Zum anderen sollten einheitliche Standards zur Messung der Daten verwendet werden. Zur Messung des Duftes eignen sich beispielsweise gaschromatographische Analysen. Durch Vortests wurde in der vorliegenden Arbeit bereits festgestellt, dass die Duftunterschiede im Genotyp 1307 (subjektiv als 'duftend' eingeordnet) und Genotyp 0744 ('nicht duftend') durch einen unterschiedlichen β -Citronellolgehalt in der Blüte bedingt sind. Die Blütenfarbe sollte entweder anhand von Messungen der farbgebenden Inhaltsstoffe, wie beispielsweise für den Anthocyangehalt (siehe Kapitel 2.3), oder zumindest anhand standardisierter Farbtafeln bei standardisierter Einstrahlung ermittelt werden.

4.2.2 Diversitätsanalyse

Die genetische Diversität spielt in einer Züchtungspopulation unter anderem eine entscheidende Rolle für das Potential zur Züchtung neuer Sorten, sowie zur Vermeidung von Inzuchtdepression. Mit einer Diversitätsanalyse ist es möglich, auf einfa-

chem Weg einen Überblick über die genetische Struktur einer Population zu erhalten. In der vorliegenden Arbeit wurde eine Diversitätsanalyse mit AFLP und dominant ausgewerteten Mikrosatellitenmarkern durchgeführt. Die Distanzen wurden nach der Methode von Nei and Li (1979) berechnet und der phylogenetische Baum wurde nach der Neighbor-Joining Methode (Saitou and Nei (1987)) konstruiert. Um die Zweige des Baumes abzusichern, wurde eine Bootstrap-Analyse mit 100 Wiederholungen durchgeführt. Ob eine ausreichende Anzahl von Markern für die Analyse verwendet wurde, kann durch Bestimmung des Variationskoeffizienten ermittelt werden (Zeid et al. (2003)). Die Autoren geben an, dass die Anzahl ausreichend ist, wenn der Mittelwert der Variationskoeffizienten unter 0,1 liegt. In der vorliegenden Studie wurden 146 Individuen mit 492 AFLP Marker und 59 dominanten Mikrosatellitenmarker getestet. Der mittlere Variationskoeffizient liegt mit 0,093 unterhalb der von Zeid et al. (2003) angegebenen Schwelle. Obwohl die Anzahl der Marker ausreichend zu sein scheint, sind viele Bootstrap-Werte relativ klein. Diese geringen Bootstrap Werte können durch die Zusammensetzung des *Phalaenopsis*-Hybriden-Genoms entstanden sein. So werden bei der Bootstrap-Methode immer ein Teil der Daten herausgenommen und durch einen anderen Teil ersetzt. Da das *Phalaenopsis*-Hybriden-Genom jedoch aufgrund der Hybridisierung und intensiven Züchtung aus mehreren Genomen (Mosaikgenom) aufgebaut ist, können kleinste Verschiebungen des Datensatzes zu unterschiedlichen Gewichtungen des jeweiligen Genomanteils führen und somit die Bootstrap-Werte verkleinern.

Bei einem Vergleich des phylogenetischen Baumes mit der Populationsstruktur ist eine große Ähnlichkeit in der Gruppierung der Genotypen zu erkennen. Wie später noch erwähnt, ist die Diversitätsanalyse eine statistisch eher ungenaue Methode zur Evaluierung der Populationsstruktur (Pritchard et al. (2000)). Allerdings wird diese Methode aufgrund der vergleichsweise einfachen und schnellen Durchführung häufig verwendet. Die Diversitätsanalyse einer Population stellt lediglich eine Momentaufnahme der aktuellen Variation dar, welche das Resultat aus einer Reihe historischer Ereignisse wie Selektion, Migration, Rekombination und der Vermehrungsart sind (Buckler and Thornsberry (2002)).

Trotz der relativ geringen statistischen Absicherung mittels Bootstrap Analyse sind Strukturen im Dendrogramm erkennbar, die die Historie der *Phalaenopsis* Züchtung widerspiegeln. Zum einen sind keine Kluster bezüglich der Herkunft der Sorten erkennbar. Das heißt, das kein Züchter Sorten hat, die sich klar von den Sorten anderer Züchter abgrenzen. Es ist vielmehr eine Vermischung der Sorten unterschiedlicher Züchter zu sehen, was darauf schließen läßt, das die Züchter gegenseitig neben dem eigenen auch den Genpool anderer Züchter nutzen.

Auf der anderen Seite gibt es eine Gruppierung in vier Hauptgruppen und meh-

rere Untergruppen. Die Hauptgruppen unterscheiden sich vor allem darin, dass in Gruppe 1 eher die älteren Sorten zu finden sind, während in Gruppe 2 neuere Sorten gefunden werden können und Gruppe 3 vorwiegend aus Wildarten besteht. Gruppe 4 unterscheidet sich vor allem in der Blütenform zu den anderen Gruppen.

Es wurde gezeigt, dass die Selektion auf positive phänotypische Merkmale einen Flaschenhalseffekt herbeiführen und so die genetische Diversität einer Population verringern kann (Buckler and Thornsberry (2002)). Dieses Phänomen wird bei *Phalaenopsis* Hybriden gerade bei der Unterscheidung von 'alten' und 'neuen' Sorten deutlich. Wie im Folgenden zu sehen, haben 'ältere' Sorten eine niedrigere Diversität als 'neuere'.

Zu Beginn der Züchtungsgeschichte von *Phalaenopsis* Hybriden waren vorrangig große, schön geformte Blüten mit einer langen Haltbarkeit das Zuchtziel. Zu diesem Zweck wurde hauptsächlich mit *Ph. amabilis* gekreuzt. Später wurden, durch die Einkreuzung von *Ph. sanderiana* und *Ph. schilleriana*, Hybriden mit rosafarbenen Blüten gezüchtet. (Gruss and Wolff (1995)) Diese 'alten' Sorten sind vorwiegend in Gruppe 1 zu finden. Aufgrund der längeren Züchtungsgeschichte ist die Diversität zwischen den Genotypen dieser Gruppe geringer als zwischen den Genotypen der anderen Gruppen. In Gruppe 1B sind vorwiegend weiße, amabilisförmige Blüten zu finden. Am Rand dieser Gruppe gibt es bereits die ersten rosa farbigen Blüten und in Gruppe 1A sind schließlich Genotypen zu finden die einen Übergang zu den rosafarbenen Hybriden bilden. Die vorherrschende Blütenform dieser Gruppe ist die Amabilisform.

Die Genotypen der Gruppen 2C und 2B sind eher zu den neueren Hybriden zu zählen. Erst in den 1960er Jahren wuchs das Interesse an farbigen und gemusterten Hybriden als Zuchtziel (Griesbach (1985)). Dies ist in diesen Gruppen an der enormen Farben- und Mustervielfalt erkennbar. Auch die Diversität ist in diesen beiden Gruppen höher als die in Gruppe 1B, welches ein weiteres Zeichen für die ältere Züchtungsgeschichte von Gruppe 1B gegenüber 2C und 2B ist. Gruppe 2B unterscheidet sich von Gruppe 2C darin, dass in Gruppe 2B fast ausschließlich pinkfarbene Blüten zu finden sind. Gruppe 2C ähnelt vom Phänotyp eher der Gruppe 2A, was darauf schließen lässt, dass auch diese Gruppe noch stark von den Wildtypen beeinflusst ist. Eine weitere Bestätigung für diese Aussage ist in den 'Bar Plots' der Analyse der Populationsstruktur zu finden. So ist in Gruppe 2C, im Gegensatz zu allen anderen untersuchten Hybriden, der höchste Anteil der Subpopulation zu finden, die auch in den Wildtypen vorhanden ist.

Die Gruppe 2A ist eine hochgradig diverse Gruppe und wird durch Gruppe 3 unterbrochen. Hier sind neuere, gelbe und gemusterte Sorten zu finden. Enorme Ähnlichkeit zwischen einigen Genotypen aus Gruppe 2A zu Genotypen aus Gruppe 3 (z.B. 1584 zu 1307) lassen darauf schließen, dass Wildtypen aus Gruppe 3 starken Einfluss auf die Züchtung dieser Sorten hatten.

Gruppe 3 schließlich wird sehr stark von Wildtypen dominiert. Im Gegensatz zu den anderen Gruppen ist hier die höchste Diversität zwischen den Genotypen zu finden. Dies zeigt deutlich, dass die genetische Diversität durch Züchtung verringert werden kann. Es zeigt aber auch, dass in der Züchtung neuer Hybriden enormes Potential steckt, wenn Wildtypen für die Kreuzungen verwendet werden.

Dieses Potential wurde offensichtlich für die Erstellung der Hybriden in Gruppe 4 genutzt. Die Blütenform der Genotypen dieser Gruppe unterscheidet sich stark von den Genotypen der anderen Gruppe. Da die Blütenform stark an *Ph. philippinensis* und *Ph. stuartiana* erinnert, und diese beiden Genotypen auch in Gruppe 4 zu finden sind, kann davon ausgegangen werden, dass diese beiden Wildtypen Ausgangsformen für die Hybriden der Gruppe 4 waren.

Eine genomweite Analyse der genetischen Variabilität in *Phalaenopsis* sp. und Hybriden wurde erst kürzlich veröffentlicht (Chang et al. (2009)). In dieser Arbeit wurden 13 *Phalaenopsis* sp. und Hybriden und drei *Doritaenopsis* sp. untersucht. Die Distanzen (ebenfalls nach der Methode von Nei and Li (1979) berechnet) liegen im Bereich von 0,298 bis 0,774. Damit sind die Distanzen in der Studie von Chang et al. (2009) höher als in der vorliegenden Arbeit. Da in der vorliegenden Arbeit hauptsächlich *Phalaenopsis* Hybriden untersucht wurden, die durch gezielte Züchtung und damit durch Selektion entstanden sind, und in der Arbeit von Chang et al. (2009) hauptsächlich unterschiedliche Arten, ist der Unterschied in den Distanzen erklärbar.

Andere molekulare Untersuchungen zur Phylogenie von *Phalaenopsis* sp. basierten nur auf Teilen des Genoms wie die ITS-Region (Tsai et al. (2005)) und das Chloroplastengenom (Padolina (2006)). Obwohl in diesen Arbeiten andere Methoden sowie andere Arten zur Baumkonstruktion verwendet wurden, sind Gemeinsamkeiten zu finden. Die beiden Arten *Ph. philippinensis* (0003) und *Ph. stuartiana* (0006) gruppieren relativ dicht zusammen und haben zu den anderen in dieser Arbeit verwendeten Wildtypen/Arten eine relativ weite Distanz. Auch die Gruppierung von *Ph. amboinensis* (0004) und *Ph. violaceae* (0001 und 0011) wurde in allen drei Bäumen verifiziert. Die Gruppierung von *Ph. pulchra* (0010), *Ph. mariae* (0009) und *Ph. hieroglyphica* (0002 und 0012) wurde ebenfalls wiedergefunden, wobei die Art *Ph. mariae* im ITS basierten Baum etwas weiter aus der Gruppe fällt. Weitere Gemeinsamkeiten konnten aufgrund unterschiedlich verwendeter Arten nicht gefunden werden.

Mit einer Distanzspanne von 0,002689 bis 0,50112 ist die Diversität in der untersuchten *Phalaenopsis* Genotypensammlung als relativ hoch einzuschätzen. Im Gegensatz dazu wurden bei anderen Zuchtpopulationen geringere Distanzen festgestellt wie z.B. in *Osteospermum* (0,000515 bis 0,101957) (Gawenda and Debener (2009)). Vergleicht man jedoch die Vermehrungsart dieser beiden Gattungen so ist die Rate der Selbstbefruchtung bei *Osteospermum* wesentlich höher als bei *Phalaenopsis*.

Buckler and Thornsberry (2002) führen auf, dass die Rate der Diversität in selbstbefruchtenden Arten häufig geringer ist, als in Fremdbefruchtern. Beispielsweise wird in der Studie von Baudry et al. (2001) an *Lycopersicum* gezeigt, dass die Diversität in selbstbefruchtenden Arten wesentlich geringer ist als in Arten, die selbstinkompatibel sind.

Der in dieser Arbeit erstellte phylogenetische Baum gibt einen Überblick über die genetische Diversität der heute auf dem europäischen Markt erhältlichen *Phalaenopsis* Hybriden. Dieser Baum kann nun genutzt werden, um die Züchtung neuer Hybriden zu verbessern, indem die Kenntnis über die genetische Distanz von potentiellen Zuchteltern in die Auswahl mit einbezogen wird.

4.2.3 Populationsstruktur

Um die Populationsstruktur zu ermitteln wurden einige Computer Programme entwickelt, die auf einer Bayesianischen Klustermethode basieren (Latch et al. (2006)). In der vorliegenden Arbeit wurde das Programm STRUCTURE 2.2 verwendet. Dieses Programm wurde gewählt, weil es von den Autoren als geeignet beschrieben wird, wenn nur wenige Informationen über die Population vorliegen und wenn 'kryptische Strukturen', die nicht sofort erkennbar sind, existieren (Pritchard et al. (2000)). Da die Populationsstruktur von *Phalaenopsis* bisher noch nicht untersucht wurde und daher keine Vorinformationen verfügbar waren, konnte mit diesem Programm eine erste Analyse der Populationsstruktur durchgeführt werden.

Da in den Ergebnissen der Berechnungen der Marker-Eigenschafts-Assoziationen zu sehen ist, dass die Populationsstruktur einen entscheidenden Einfluss auf das Ergebnis hat, wurde in der vorliegenden Arbeit besonders viel Wert auf diese Analyse gelegt. Der BurnIn und die Wiederholungen wurden in STRUCTURE 2.2 möglichst hoch gesetzt, es wurden vier unabhängige Durchläufe ausgeführt und die Anzahl der Subpopulationen wurde mit drei unterschiedlichen Methoden berechnet. Damit konnte, trotz der relativ kleinen Anzahl an Markerdaten, die Anzahl der Subpopulationen relativ genau bestimmt und durch mehrere Wiederholungen und unabhängige Vergleiche verifiziert werden.

Um die Anzahl der Subpopulationen und die Zugehörigkeit der Genotypen zu den jeweiligen Subpopulationen zu bestimmen, wurden drei Methoden genutzt. Bevor die Anzahl der Subpopulationen jedoch ermittelt werden kann, ist eine Definition der (Sub-) Populationen wichtig. Waples and Gaggiotti (2006) zeigen auf, daß es zwei Extreme geben kann. Bei isolierten Populationen sind die Genotypen der einzelnen Population vollständig voneinander getrennt. Das heißt aufgrund von zum Beispiel zeitlicher oder räumlicher Trennung gibt es keinen Genfluss zwischen den einzel-

nen Populationen. Das andere Extrem hierzu ist die Panmixie, bei der alle Genotypen/Individuen sich mit gleich hoher Wahrscheinlichkeit paaren können. Zwischen diesen beiden Extremen gibt es jedoch viele Abstufungen. Bei den hier untersuchten *Phalaenopsis* Hybriden ist auf der einen Seite eine starke Abgrenzung bzw. Isolierung zu den Wildarten aufgrund der großen geographischen Distanz gegeben. Wie bereits in Kapitel 1.1 beschrieben, gehen die meisten Hybriden auf nur wenige Wildarten zurück. Die geographische Abgrenzung ist besonders bei den 'alten' Sorten/Hybriden (Gruppe 1B) deutlich. In Abbildung 3.19 ist zu sehen, dass diese Sorten mit einer relativ hohen Wahrscheinlichkeit in eine Subpopulation eingeteilt werden. Im Gegensatz dazu ist die Einteilung 'neuerer' Sorten/Hybriden (z.B. Gruppe 2C) weniger eindeutig. Hier wird deutlich, daß es aufgrund der Züchtung mit erst in neuerer Zeit eingeführter Wildarten und Hybriden bzw. durch die Nutzung des gesamten im europäischen und asiatischen Raum verfügbaren Genpools zu einer Durchmischung der Populationen (Admixture) kam. Wie bei den meisten realen Populationen (Waples and Gaggiotti (2006)) befindet sich auch die hier untersuchte Population zwischen den beiden oben genannten Extremen. Während die Methode von Pritchard et al. (2000) und die Methode von Cockram et al. (2008) von Populationen ausgehen, deren Subpopulationen gut voneinander abgegrenzt sind, wird in der Methode von Evanno et al. (2005) die Möglichkeit des Genflusses zwischen einzelnen Subpopulationen mit betrachtet.

Pritchard et al. (2000) beschreiben ihre Methode selbst als eine *ad hoc* Annäherung an die wahre Populationsstruktur und räumen ein, dass eine biologische Interpretation der Ergebnisse schwierig sein kann. Ihre Methode funktioniert gut für kleine Datensätze mit gut abgegrenzten Populationen. Dies ist im biologischen Sinn jedoch eher selten bis gar nicht anzutreffen und daher bietet die Methode lediglich eine gute Annäherung an die wahre Populationsstruktur. In vier unabhängigen Durchläufen wurden zweimal zehn, einmal acht und einmal zwölf Subpopulationen berechnet (siehe Tabelle 3.8). Da die Methode von Pritchard et al. (2000) auf Grundlage idealisierter Daten entwickelt wurde, ist sie nur mit Vorsicht auf reale Daten anwendbar. Die Autoren zeigen auf, dass die Anwendung eine grobe Einschätzung der Anzahl der Subpopulationen ist. Daher sind die unterschiedlichen Ergebnisse der vier Durchläufe als Annäherung an die Anzahl der vorhandenen Subpopulationen zu bewerten.

Bei der Methode nach Cockram et al. (2008) werden die LnP(D) -Werte von allen vier Durchläufen in einer Graphik dargestellt und können somit gleichzeitig betrachtet werden. Es ist zu sehen, daß die LnP(D) -Werte einem Trend folgen. Allerdings können einzelne Rechnungen unter Umständen stark von diesem Trend abweichen. In der Methode nach Pritchard wird jeder Durchlauf einzeln betrachtet und folglich können so unterschiedliche Ergebnisse in der Anzahl der Subpopulationen errechnet werden. In der Methode nach Cockram können, durch die gleichzeitige Betrachtung

aller Durchläufe, Ausreißer als solche erkannt werden. Der Einfluss der Ausreißer wird dadurch minimiert. Durch die Abschätzung der Korrelationen zwischen den Q-Matrizen kann die Anzahl an Subpopulationen relativ genau abgeschätzt werden. Obwohl die graphische Methode nicht die Genauigkeit der zuvor beschriebenen Methode hat, so ist diese bei realen Daten jedoch vorzuziehen. Denn aufgrund der globalen Sicht auf alle Durchläufe können auf diese Weise Fehler bzw. Ausreißer neutralisiert werden. Die beiden hier beschriebenen Methoden sind sich sehr ähnlich und beruhen beide auf der Annahme das die Subpopulationen gut voneinander abgegrenzt sind.

Evanno et al. (2005) versuchen einen mathematischen Ansatz, um die Anzahl an Subpopulationen zu ermitteln. Dabei werden die $\ln P(D)$ -Werte (bei Evanno als $L(K)$ bezeichnet) wieder in Abhängigkeit zu der vorgegebenen Anzahl an Subpopulationen gesetzt. Anschließend wurde durch Bildung der ersten und zweiten Ableitung (nach den in Kapitel 2.5.6.3 genannten Formeln) ΔK bestimmt. Durch Dividieren der Werte mit der Standardabweichung konnten die Varianzen ebenfalls in die Berechnung mit einfließen. Die Grundidee der Ermittlung von ΔK ist es, den Punkt deutlich zu machen, der die Kehrtwende in der Verteilung der $\ln P(D)$ -Werte in Abhängigkeit der Anzahl der Subpopulation zeigt. Die Autoren gehen davon aus, das sich an diesem Punkt die 'wahre' Anzahl der Subpopulationen befindet. Im Gegensatz zu den beiden zuvor genannten Methoden werden die Subpopulationen hier nicht als getrennte Populationen betrachtet. In allen drei bei Evanno et al. (2005) zugrunde liegenden Modellen (Kontaktzone \rightarrow Genfluss zwischen jeweils zwei Subpopulationen, Inselmodel \rightarrow Genfluss zwischen allen Subpopulationen, hierarisches Inselmodel \rightarrow Genfluss innerhalb und zwischen einem Set von Subpopulationen) wird Genfluss zwischen den Subpopulationen mitbetrachtet. Demnach ist diese Methode grundsätzlich für Populationen geeignet, bei denen eine Durchmischung des Genpools angenommen wird.

Da sich die Ergebnisse der drei hier angewandten Methoden sehr ähneln und auch eine Übereinstimmung der Subpopulationen mit der Gruppierung der Diversitätsanalyse zu beobachten ist, ist eine eindeutige Aussage bezüglich der richtigen oder falschen Wahl einer Methode nicht zu treffen.

4.2.4 Vergleich Diversitätsanalyse und Ermittlung der Populationsstruktur

Pritchard et al. (2000) zeigen einige Nachteile von Distanz-basierten Methoden wie beispielsweise Neighbour-Joining zur Detektion der Populationsstruktur auf. Zum einen ist die Gruppierung der Individuen stark von der verwendeten Distanzmethode und von der Darstellungsweise abhängig. Es ist schwer abzuschätzen wie gut die Aussagekraft der berechneten Gruppen ist. Des Weiteren ist es schwierig zusätzliche

Information, wie bereits bekannte Gruppierungen z.B. durch gleiche Herkünfte, in die Berechnung mit einfließen zu lassen. Daher bezeichnen Pritchard et al. (2000) Ergebnisse auf Grundlage von Distanz-basierten Methoden eher informativ als eine exakte statistische Schlussfolgerung. Dies wird auch bei der Gruppierung der Genotypen deutlich. Während auf Basis des Dendrogramms nur acht Gruppen mehr oder weniger deutlich von einander abgegrenzt werden konnten, wurde anhand der Berechnungen der Populationsstruktur mit STRUCTURE 2.2 zwischen zehn und zwölf Gruppen/Subpopulationen validiert. Dies zeigt die wesentlich bessere Auflösung einer solchen Berechnung.

In Abbildung 3.19 ist die Populationsstruktur neben dem Neighbour-Joining Dendrogramm dargestellt. Während im Dendrogramm die Genotypen 5700, 0738 und 0731 eine Untergruppe bilden und die Genotypen *Phalaenopsis phillipiensis* (0003), 5702 und *Phalaenopsis stuartiana* (0006) eine andere Untergruppe, haben beide Gruppen in der Populationsstruktur (berechnet mit STRUCTURE 2.2 und der Annahme K=10 Subpopulationen) einen sehr ähnlichen genetischen Hintergrund, welcher im Dendrogramm nicht deutlich wird. So beinhalten die Genotypen der zuerst genannten Gruppe das genetische Material aus zwei Subpopulationen, aus denen die Genotypen der zuletzt genannten Gruppe größtenteils bestehen (farblich in blau und orange dargestellt). Damit sind die Ergebnisse der Populationsstruktur, die mit STRUCTURE 2.2 und damit auf Grundlage einer modelbasierten Methode errechnet wurden, eher für die Berechnung von Assoziationen geeignet als Ergebnisse, die auf distanzbasierten Methoden beruhen.

4.2.5 Kopplungsungleichgewicht

Eine weitere Voraussetzung für die Durchführung und Interpretation der Ergebnisse einer Assoziationsstudie ist die Kenntnis über das Kopplungsungleichgewicht der zu untersuchenden Population. In der vorliegenden Arbeit wurde das Kopplungsungleichgewicht zwischen den für die Assoziationsstudie genutzten Markern mit dem Programm TASSEL 2.1 ermittelt. Um einen Eindruck über das allgemeine Verhalten des Kopplungsungleichgewichtes in einer *Phalaenopsis* Population zu erhalten, wurde das Kopplungsungleichgewicht zwischen den kartierten Markern ebenfalls mit TASSEL 2.1 ermittelt und in Abhängigkeit zum genetischen Abstand in cM graphisch dargestellt (siehe Abbildung 3.23).

Bis zu diesem Zeitpunkt wurden noch keine Arbeiten veröffentlicht, in der das Kopplungsungleichgewicht in *Phalaenopsis* oder einer anderen Gattung der Familie *Orchidaceae* untersucht wurde.

In Abbildung 3.23 ist das Kopplungsungleichgewicht der kartierten Marker in Ab-

hängigkeit zum Abstand aufgetragen. Innerhalb von 30 cM sinkt das Kopplungsungleichgewicht um die Hälfte von eins auf 0,5 ab. Es wurde gezeigt, dass die Rekombinationsrate bei Selbstbefruchtern wesentlich geringer ist als bei Fremdbefruchtern, da einzelne Loci bei Selbstbefruchtern eher homozygot sind als bei Fremdbefruchtern (Nordborg and Donnelly (1997), Flint-Garcia et al. (2003)). Wie in Abbildung 1.2 zu sehen ist, sinkt das Kopplungsungleichgewicht bei einer höheren Rekombinationswahrscheinlichkeit schneller ab, als bei Populationen mit niedriger Rekombinationswahrscheinlichkeit. *Phalaenopsis* sp. vermehren sich fast ausschließlich durch Fremdbefruchtung. Dies lässt vermuten, daß die Rekombinationswahrscheinlichkeit relativ hoch ist und demzufolge ist ein relativ schneller Abfall des Kopplungsungleichgewichts in Abhängigkeit des Abstands zu erwarten.

Da die Abnahme des Kopplungsungleichgewichts in fast allen Studien im Vergleich zum physikalischen Abstand der Marker untersucht wurde, aber die Abstände der Marker im *Phalaenopsis* Genom noch nicht auf physikalischer Ebene geklärt werden konnten, ist ein Vergleich mit anderen Studien schwierig. Flint-Garcia et al. (2003) zeigen als Beispiel, dass das Kopplungsungleichgewicht bei Fremdbefruchtern innerhalb von 500 bp zerfällt, während es bei hochgradig selbstbefruchtenden Populationen durchaus über 10 kb (bei *Arabidopsis* bis zu 250 kb [Nordborg and Innan (2002)]) erhalten bleiben kann. Gupta et al. (2005) fassen einige Arbeiten zusammen, die starke Schwankungen in der Geschwindigkeit des Abfalls des Kopplungsungleichgewichts verschiedener Arten zeigen. Gerste und tetraploider Weizen zeigen einen Abfall innerhalb von 10-20 cM, während bei Hirse (ein weiterer Selbstbefruchter) der Abfall des Kopplungsungleichgewichts innerhalb von 4 cM erfolgt. Fremdbefruchter wie Zuckerrohr und Zuckerrübe zeigen einen Abfall des Kopplungsungleichgewichts innerhalb von 10 bzw. 3 cM (Studien zitiert in Gupta et al. (2005)). Demzufolge ist ein Abfall des Kopplungsungleichgewichts innerhalb von 30 cM als sehr langsam zu bewerten.

In Tabelle 4.1 sind einige Faktoren aufgeführt, die das Kopplungsungleichgewicht (LD) beeinflussen. Unter anderem wird das Kopplungsungleichgewicht durch die Größe der Population, Selektion und genetische Isolierung beeinflusst.

Schaut man sich nun die Züchtungsgeschichte der untersuchten *Phalaenopsis* Hybriden genauer an, so ist zu sehen, daß die heutigen Hybriden auf nur wenige Genotypen zurückzuführen sind. Gruss and Wolff (1995) zeigen auf, dass viele Hybriden einen neuen Namen erhielten und somit irrtümlich als neue, eigenständige Sorte betrachtet werden. Genetisch gehen die heutigen Hybriden größtenteils auf wenige Wildarten wie *Ph. aphrodite*, *Ph. equestris*, *Ph. amabilis*, *Ph. violaceae*, *Ph. schilleriana* und *Ph. sanderiana* zurück.

Des Weiteren führen Gruss and Wolff (1995) auf, dass zu Beginn der Hybridisierungsgeschichte vorwiegend auf große, schön geformte Blüten selektiert wurde und zu

Tabelle 4.1: Faktoren, die das Kopplungsungleichgewicht (LD) in einer Population beeinflussen (aus: Rafalski and Morgante (2004) und Oraguzie et al. (2006))

Einflussfaktor	Effekt
Rekombinationsrate	höhere Rekombination verringert LD
Vermehrung durch Selbstbefruchtung	hohes LD
Vermehrung durch Fremdbefruchtung	niedriges LD
genetische Isolierung zwischen Linien	Erhöhung des LD
Population Subdivision	Erhöhung des LD
Vermischung der Populationen	Erhöhung des LD
natürliche und künstliche Selektion	lokale Erhöhung des LD
Populationsgröße	kleine Populationen haben mehr LD
Balancierte Selektion	Erhöhung des LD
Mutationsrate	hohe Mutationsrate verringert das Gesamt-LD, allerdings ist das lokale LD um die neue Mutation sehr hoch, bis es durch Rekombination wieder verringert wird
Genetische Umordnung	durch genetische Umordnung wird die lokale Rekombination unterdrückt, was zu einer Erhöhung des LD in dieser Umgebung führt
Stochastische Effekte	Erhöhung oder Verringerung des LD
Epistatische Effekte	Erhöhung des LD

diesem Zweck vorwiegend mit *Ph. amabilis* gekreuzt wurde. Auch heute noch ist die so genannte amabilisförmige Blüte ein positives Selektionskriterium. Eine Selektion für oder gegen ein bestimmtes Merkmal kann zu einer Erhöhung des Kopplungsungleichgewichts zwischen den, für das Merkmal verantwortlichen Genen, führen, obwohl keine physikalische Kopplung zwischen diesen Genen vorliegt (epistatische Effekte) (Sorkheh et al. (2008)). Daher ist das Kopplungsungleichgewicht um dieses Merkmal herum vermutlich erhöht.

Da bei *Phalaenopsis* nicht nur auf Blütengröße sondern gleichzeitig auch auf andere Merkmale, wie z.B. Haltbarkeit der Blüten, gezüchtet wurde, haben diese Merkmale eine positive Selektion durchlaufen. Durch diese balancierte Selektion mehrerer Merkmale hat sich das Kopplungsungleichgewicht zwischen den Genen, die für die entsprechenden Merkmale kodieren, erhöht.

Eine genetische Isolierung der europäischen Kulturhybriden ist durch die große geographische Distanz zum ursprünglichen Verbreitungsgebiet der *Phalaenopsis* Wildarten gegeben. Während das ursprüngliche Verbreitungsgebiet sich im südöstlichen Teil Asiens und Ozeanien befindet (siehe Abbildung 1.1) wurden die ersten Kulturhybriden vorwiegend in England und Frankreich von nur wenigen Züchtern entwickelt (Gruss and Wolff (1995)).

Wie Zhu et al. (2008) beschreiben, ist die Auflösung in einer Assoziationskartie-

rungsstudie relativ hoch, wenn das Kopplungsungleichgewicht innerhalb kurzer Distanzen abnimmt. Allerdings wird dann auch eine relativ hohe Anzahl an Markern benötigt. Anders herum ist bei einem langsamen Abfall des Kopplungsungleichgewichts nur eine relativ geringe Anzahl von Markern erforderlich, jedoch ist dann nur eine grobe Analyse aufgrund der schlechteren Auflösung möglich. Ein weiterer Faktor, der zur Schätzung der Anzahl an benötigten Markern für eine Assoziationsstudie beachtet werden muss, ist die Tatsache, daß genauso wie die Rekombinationsfrequenz (niedriger in der Nähe des Zentromers) auch das Kopplungsungleichgewicht im Genom variiert. Daher ist das lokale Kopplungsungleichgewicht und damit auch die Anzahl der notwendigen Marker für eine Analyse schwer zu bestimmen (Sorkheh et al. (2008)). Zur Zeit existieren keine Veröffentlichungen, die eine ungefähre Anzahl der benötigten Marker nennen.

Da das Kopplungsungleichgewicht in der untersuchten Population relativ langsam abnimmt und, wie oben beschrieben, in diesem Fall nur eine relativ kleine Anzahl von Markern nötig ist, ist davon auszugehen, dass eine relativ grobe Assoziation der Marker mit dem zu untersuchenden Merkmal in dieser Arbeit erreicht werden kann. Da bisher keine Arbeiten zur genetischen Struktur von *Phalaenopsis* Hybriden veröffentlicht wurden, kann diese grobe Assoziation die Grundlage für weitere Untersuchungen und Feinkartierung von interessierenden Merkmalen bilden.

4.2.6 Berechnung der Marker-Eigenschafts-Assoziationen

Nachdem alle Voraussetzungen für eine Assoziationsstudie betrachtet wurden, konnten die Marker-Eigenschafts-Assoziationen berechnet werden. Dazu wurden zwei Ansätze genutzt. Zum einen wurde, wie in Achleitner et al. (2008) beschrieben, ein GLM Ansatz mit der Populationsstruktur als Kofaktor in TASSEL 2.1 gerechnet. Im zweiten Ansatz wurde eine Methode gewählt, die ebenfalls auf GLM beruht und neben der Populationsstruktur eine 'Genomic Control' Korrektur durchführt. Dieser Ansatz wurde mit einer, von J. White (2009) geschriebenen Routine (siehe 5.6) in R 2.8.1 gerechnet.

Beiden Methoden liegt zugrunde, dass es sich um Single Marker Eigenschafts Assoziationen handelt. Aufgrund der geringen Markerdichte wurden die Marker jeweils einzeln auf eine Assoziation mit der zu untersuchenden Eigenschaft getestet. Eventuelle Zusammenhänge oder Einflüsse anderer Marker/-regionen wurden nicht betrachtet. Dies ist bei qualitativen Eigenschaften, welche meist durch einzelne Gene reguliert werden ein guter Ansatz, welcher in mehreren Studie bereits erfolgreich durchgeführt wurde (z.B. Hansen et al. (2001), D'hoop et al. (2008)). Jedoch werden quantitative Eigenschaften meist durch mehrere Gene beeinflusst, die unter Umständen nur einen

kleinen Effekt haben können, und deshalb schwerer zu detektieren sind als Gene die für eine qualitative Eigenschaft kodieren (Besenbacher et al. (2008)). Aus diesem Grund können in der vorliegenden Studie für quantitative Eigenschaften wie Blütengröße und Rispenlänge lediglich Marker gefunden werden, die Hinweise auf QTL geben, welche unter Umständen nur zum Teil für das entsprechende Merkmal verantwortlich sind. Allerdings bietet die Assoziationsstudie das Potential, diese QTL bei einer höheren Markerdichte und entsprechender statistischer Verfahren, welche das Kopplungsungleichgewicht flankierender Regionen mit betrachten, relativ genau zu verifizieren und zu lokalisieren (z.B. Brown et al. (2008), Salvi et al. (2007)).

Das Kopplungsungleichgewicht zwischen den Markern ist mit unter 1% sehr klein. Aufgrund des geringen Kopplungsungleichgewichts zwischen den Markern ist davon auszugehen, daß es bei den Berechnungen der Marker-Eigenschafts-Assoziationen keine falsch positiven Ergebnisse aufgrund des Kopplungsungleichgewichts zwischen den Markern gab. Eine Berechnung der einzelnen Assoziationen pro Marker ist hier auch von Vorteil, da die Marker sich vermutlich aufgrund des geringen Kopplungsungleichgewichts kaum beeinflussen. Ein Nachteil des geringen Kopplungsungleichgewichts zwischen den Markern ist, das Assoziationen nur schwer zu detektieren sind. Des Weiteren können diese Assoziationen nicht durch umliegende Marker verifiziert werden. Da nur wenige Marker im Kopplungsungleichgewicht gefunden wurden, ist davon auszugehen, daß die Genomabdeckung mit den vorhandenen Markern nur rudimentär ist.

In TASSEL 2.1 wurde ein GLM Ansatz mit der Populationsstruktur aus Structure als Kofaktor durchgeführt. Als statistische Absicherung wurde ein Permutationstest mit 1000 Wiederholungen durchgeführt. Als p-Wert wurde zum einen ein nichtadjustierter Wert von 0,05 und ein bonferoniadjustierter p-Wert von $0,05/492$ angenommen. In R 2.8.1 wurde ebenfalls die Populationstruktur als Kofaktor in das Modell mit einbezogen. Hier wurde jedoch eine 'Genomic Control' (GC) Korrektur durchgeführt. Bei der GC-Korrektur werden zufällige Marker genutzt, um die Signifikanz der Teststatistik zu erhöhen, indem davon ausgegangen wird, das eine vorhandene Populationsstruktur alle Genorte gleichermaßen betrifft (Yu et al. (2006)). Als Signifikanzschwelle wurde ein bonferoniadjustierter p-Wert von $0,05/492$ angenommen. Durch die 'Genomic Control' Korrektur wurden die berechneten p-Werte kleiner. Dadurch wurde es schwieriger, das bonferoniadjustierte Signifikanzniveau zu erreichen und der Test wurde somit konservativer. Im Fall der Blütengröße konnten so ohne Beachtung der Populationsstruktur keine signifikanten Marker mehr detektiert werden. Mackay and Powell (2007) präferieren bei Studien mit geringer Markerdichte die 'Genomic Control' Korrektur trotz der konservativeren Testmethode. Sie fügen an, daß auf lange Sicht falsch negative Ergebnisse weniger problematisch sind als falsch

positive Ergebnisse.

Legt man den bonferoniadjustierten p-Wert von 0,05/492 zu Grunde, so wurden mit der in TASSEL 2.1 berechneten Methode wesentlich weniger assoziierte Marker gefunden als in der in R 2.8.1 berechneten Methode. Dies lässt vermuten, dass die zuerst genannte Methode noch konservativer ist als die zuletzt genannte.

In der vorliegenden Arbeit wurden zwei statistische Methoden zur Detektion von Marker-Eigenschafts-Assoziationen durchgeführt. Obwohl im Gegensatz zu konventionellen Kartierungsmethoden die Entwicklung robuster Methoden für die Assoziationskartierung in Pflanzen bisher vernachlässigt wurde (Zhu et al. (2008)), gibt es weitere Ansätze zur Berechnung dieser Assoziationen. Yu et al. (2006) stellen ein Modell vor, welches neben der Populationsstruktur auch die familiäre Struktur mit einbezieht. Price et al. (2006) beziehen hingegen die Populationsstruktur als 'Principal Component Analysis' (PCA) Korrektur ein. In weiterführenden Studien kann getestet werden, ob diese Methoden einen Vorteil gegenüber den hier angewandten Methoden haben.

Bei der Berechnung von Assoziationen zwischen Marker und Eigenschaften wurden Marker gefunden, die scheinbar einen Einfluss auf mehrere Eigenschaften haben. Dies lässt vermuten, daß diese Marker an Genloci gekoppelt sind, die einen pleiotropen Einfluss haben. Ob es sich hierbei jedoch um 'echte' Assoziationen handelt, oder um falsch positive Ergebnisse, bleibt durch weitere Analysen wie beispielsweise durch Klonierung der QTL, zu überprüfen.

Nicht alle Marker, die eine signifikante Assoziation zu den untersuchten Eigenschaften zeigten, konnten kartiert werden. Sie waren in der Kartierungspopulation nicht polymorph. Um die entsprechenden Marker zu kartieren, sollte eine Nachkommenschaft erzeugt werden, die in dem entsprechenden Merkmal phänotypisch wie auch genotypisch aufspaltet.

Aufgrund der Tatsache, daß keine geeignete Population für eine familienbasierte QTL-Analyse zur Verfügung stand, bietet die Methode der genomweiten Assoziationsstudie einen guten Ansatz zur Detektion von Marker-Eigenschafts-Assoziationen. Allerdings hängt eine solche Studie stark von der Anzahl und Qualität der genetischen Marker und phänotypischen Merkmale ab. Während in vielen anderen Studien hunderttausende von molekularen Markern, welche über das gesamte Genom verteilt sind, genutzt wurden (Zhu et al. (2008)), wurde hier mit nur wenigen Markern für ein relativ großes Genom, ein erster Schritt zur Analyse des Genoms gemacht, welcher als Grundlage für weitere Studien dienen kann. Allerdings zeigen die durchgeführten Analysen, daß die Anwendung genomweiter Assoziationsstudien für die Detektion von Marker-Eigenschafts-Assoziationen in *Phalaenopsis* gut geeignet sind. Eine Verifizierung der detektierten QTL kann nun durch Erstellung einer für die entsprechende Eigenschaft spezifischen Nachkommenschaft erreicht werden. Des Weiteren können

die QTL durch Klonierung und Sequenzierung genauer spezifiziert werden. Sobald die QTL verifiziert wurden, können sie für die Züchtung neuer, verbesserter Hybriden durch markergestützte Selektion verwendet werden.

4.3 Vergleich familienbasierte QTL-Analyse und genomweite Assoziationsstudie

In der vorliegenden Studie wurde eine familienbasierte QTL-Analyse und eine genomweite Assoziationsstudie an *Phalaenopsis* Hybriden durchgeführt. Zu Beginn des Projektes war nur eine Nachkommenschaft der Eltern 19 und 21 (siehe Abschnitt 2.1) verfügbar. Es wurde aufgrund der geringen phänotypischen Unterschiede der Eltern, nicht erwartet eine zufriedenstellende Segregation von gartenbaulich interessanten Merkmalen zu beobachten. Aus diesem Grund wurde eine Assoziationsstudie in einer Population, die eine Stichprobe des heute auf dem europäischen Markt erhältlichen *Phalaenopsis* Hybriden Sortiments entspricht, angestrebt.

Der familienbasierten QTL-Analyse sowie der genomweiten Assoziationsstudie liegt zu Grunde, dass aufgrund des Kopplungsungleichgewichts die flankierenden DNA Regionen eines merkmalausprägenden DNA Abschnitts mitvererbt werden (Zhu et al. (2008)). Die Vor- und Nachteile beider Methoden wurde bereits in mehreren Veröffentlichungen diskutiert (z.B. Mackay and Powell (2007), Nordborg and Weigel (2008), Gupta et al. (2005), Abdurakhmonov and Abdukarimov (2008), Zhu et al. (2008)) und werden im Folgenden zusammengefasst.

In der genomweiten Assoziationsstudie werden historische Rekombinationsereignisse mitbetrachtet, während bei der familienbasierten QTL-Analyse nur aktuelle Rekombinationsereignisse erfasst werden können. So muss die Nachkommenschaft groß genug sein, um QTL möglichst präzise zu detektieren.

Die Möglichkeit bestimmte QTL gezielt zu detektieren ist in einer familienbasierten QTL-Analyse größer als in einer genomweiten Assoziationsstudie, da gezielte Kreuzungen auf das bestimmte Merkmal durchgeführt werden (Mackay and Powell (2007)). So können hier auch seltene Allele, die unter Umständen einen großen phänotypischen Effekt haben, detektiert werden.

QTL, welche in einer spezifischen Kreuzung entstanden sind, können jedoch unter Umständen auf die spezifische Kreuzung und deren genetischen Hintergrund zurück zu führen sein. Deshalb kann es schwierig sein, diese QTL auf andere Populationen zu übertragen (Achleitner et al. (2008)). Es besteht die Möglichkeit, daß die Allele nur in der untersuchten Nachkommenschaft für die Ausprägung des entsprechenden Merkmals verantwortlich sind. Diese seltenen Allele, welche einen großen Effekt haben

können, werden bei der genomweiten Assoziationsanalyse leichter verfehlt. Andererseits werden hier eher allgemeine Allele mit eventuell kleineren Effekten detektiert, welche in einer Kartierungspopulation eventuell nicht vorhanden sind. In einer spezifischen Nachkommenschaft, können diese allgemeinen Allele nur geringen Einfluss nehmen. Das heißt, Allele die in einer genomweiten Assoziationsstudie detektiert wurden, können eher auf die gesamte Gattung übertragen werden, als solche, die in einer familienbasierten QTL-Analyse für eine bestimmte Population detektiert wurden.

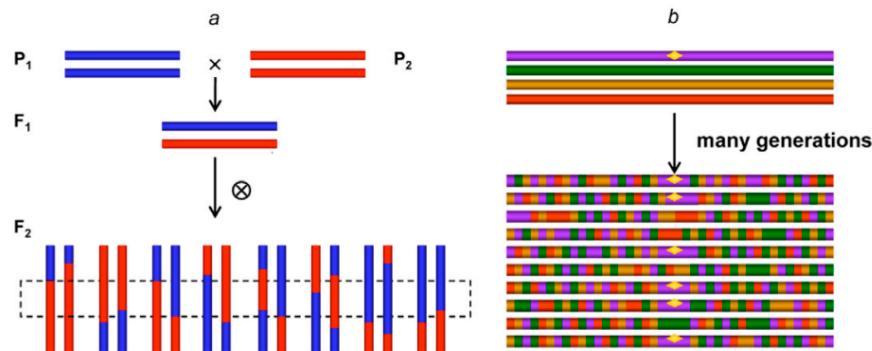


Abbildung 4.1: Schematischer Vergleich zwischen familienbasierte Kopplungsanalyse und Assoziationsstudie mit einer Kollektion von verschiedenen Genotypen (aus: Zhu et al. (2008)): In der familienbasierten Kopplungsanalyse (a) gibt es nur wenige Rekombinationsereignisse, was zu einer geringeren Auflösung in der Kartierung führt. Hingegen kann, aufgrund vieler Rekombinationsereignisse, bei der Assoziationsstudie (b) eine höhere Auflösung erzielt werden.

In Abbildung 4.1 sind schematisch die mögliche Anzahl von Rekombinationen um einen bestimmten Locus, in einer Nachkommenschaft zweier Eltern (a) und einer Population mehrerer Vorfahren und Generationen (b), dargestellt. Es ist zu sehen, dass in einer familienbasierten Studie mit einer experimentellen Nachkommenschaft, nur wenige Rekombinationsereignisse geschehen können. In einer Assoziationsstudie werden hingegen mehrere historische Rekombinationsereignisse sowie die natürliche Diversität ausgenutzt, was in einer höheren Auflösung in der Kartierung von Merkmalen resultiert (Zhu et al. (2008)). Damit ist in einer Assoziationskartierung eine wesentlich feinere Kartierung von Merkmalen möglich. Beispielsweise ist laut Remington et al. (2001) in *Zea mays* eine bis zu 5000fach höhere Auflösung in der Kartierung von QTL möglich als in F_2 oder rekombinanten Inzucht Populationen.

Eine Möglichkeit die Vorteile beider Kartierungsmethoden auszunutzen bietet die so genannte 'Nested-Association-Mapping' (NAM) Methode, in der die hohe Auflösung der Assoziationsstudie mit der Robustheit der familienbasierten Studien verbunden wird (Nordborg and Weigel (2008)).

4.4 Ausblick

In der vorliegenden Arbeit wurde versucht einen Überblick über die genomische Struktur von *Phalaenopsis* Hybriden zu erhalten und Marker bzw. Genregionen zu finden, die die Züchtung neuer Sorten durch Markergestützte Selektion verbessern bzw. beschleunigen können. Allerdings ist bisher kaum etwas über das *Phalaenopsis*-Genom bekannt. In Abbildung 4.2 sind die Schritte aufgeführt, welche notwendig sind, um die Markergestützte Selektion in einer Gattung zu etablieren. Die vorliegende Arbeit befasst sich mit den ersten beiden Punkten. Es wurde eine erste genetische Karte erstellt und QTL mit zwei verschiedenen Ansätzen gefunden.

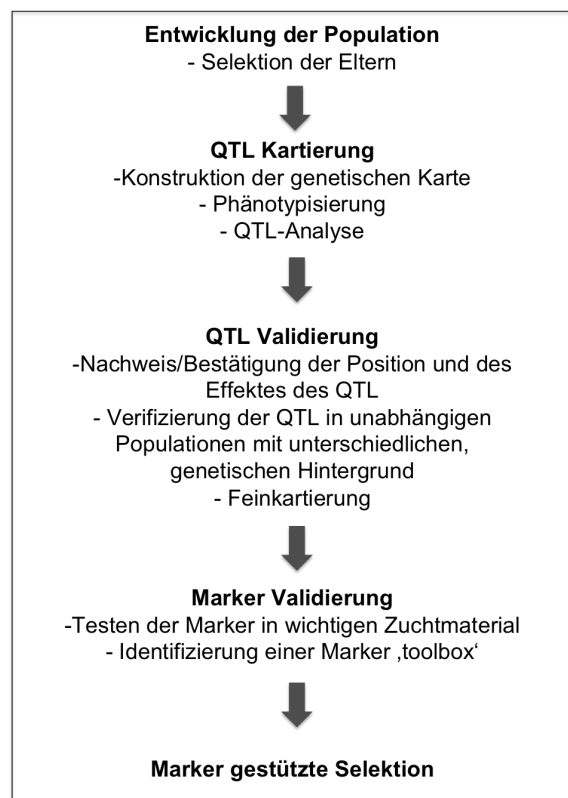


Abbildung 4.2: Übersicht der Markerentwicklung für die Markergestützte Selektion (nach: Collard and Mackill (2008))

Gene oder QTL in einem Genom, insbesondere in einem solch großen und komplexen Genom wie dem von *Phalaenopsis* zu finden, ist ähnlich wie die Nadel im Heuhaufen zu suchen (Collard et al. (2005)). In dieser Arbeit war es bereits möglich Marker bzw. Regionen zu identifizieren, die einen Einfluss auf das entsprechende Merkmal haben können. Der nächste Schritt ist nun diese QTL zu verifizieren. Um die Position und den phänotypischen Effekt eines QTL zu verifizieren, muss dieser in unabhängigen Populationen mit unterschiedlichem genetischen Hintergrund unter-

sucht werden (Collard and Mackill (2008)). Idealerweise sollte eine neue *Phalaenopsis* Nachkommenschaft erzeugt werden, welche sich deutlicher in den zu untersuchenden Eigenschaften unterscheidet, als die in der vorliegenden Arbeit verwendete Nachkommenschaft. Des Weiteren sollten Klone gleicher Genotypen in unterschiedlichen kontrollierten Umwelten über einen längeren Zeitraum phänotypisiert werden, um Fehler bei der Berechnung der QTL zu verringern/vermeiden. Um eine genauere Position der QTL zu erreichen, muss die genetische Karte eine höhere Abdeckung des Genoms leisten. Dazu müssen mehr Marker in einer größeren Nachkommenschaft in die Karte integriert werden.

Erst nachdem die QTL und anschließend die entsprechenden Marker verifiziert wurden, können diese für eine Markergestützte Züchtung von *Phalaenopsis* Hybriden eingesetzt werden. Obwohl die Verifizierung der QTL und Marker noch einen hohen Arbeits- und Zeitaufwand verlangt, so ist dies gerade bei *Phalaenopsis* Hybriden eine sinnvolle Investition. Zur Zeit basiert die klassische Züchtung von neuen Hybriden vorwiegend auf den Erfahrungen des Züchters und es gibt keine genetisch basierte Zuchtstrategie. Die Auswahl der Eltern erfolgt aus bekannten Genotypen deren Eigenschaften der Züchter aus früheren Kreuzungen oder der Herkunft kennt. Dabei weiß der Züchter von vielen Eigenschaften, ob sie maternal oder paternal vererbt werden und entscheidet daraufhin welcher Genotyp die Mutter- bzw. Vaterpflanze wird (F. von Rundstedt, Bock Bio Science, persönliche Mitteilung). In der F1-Generation wird schließlich auf Ausprägung der Merkmale selektiert. Diese meist subjektive und langwierige Züchtungsstrategie kann durch Markergestützte Selektion verbessert werden. Die drei wichtigsten Vorteile der Markergestützten Selektion (MAS) werden von Collard and Mackill (2008) genannt:

Ein Markerscreening kann einfacher sein als die Phänotypisierung der Nachkommen. So ist die Aufnahme von Merkmalen die den Zierwert der Pflanze betreffen oft subjektiv, wie beispielsweise die Farbausprägung oder der Blütenduft. Meist ist eine objektive Aufnahme der Daten mit großem Aufwand und mehreren Messungen verbunden.

Die Selektion der gewünschten Genotypen kann bereits im Saatbeet erfolgen. Gerade bei *Phalaenopsis* Hybriden ist dieser Punkt einer der wichtigsten. Vom Zeitpunkt der Aussaat bis zur Blüte können mehrere Jahre vergehen. Findet eine Selektion bereits im Saatbeet statt, kann die Nachkommenschaft auf wenige Genotypen reduziert werden. Dies bedeutet eine große Platz- und Zeitersparnis, da nur noch wenige Genotypen kultiviert und bewertet werden müssen. Dadurch ergibt sich die Möglichkeit, mehr Kreuzungen im gleichen Zeitraum durchzuführen.

Einzelne Pflanzen können selektiert werden. Um Pflanzen in ihren Eigenschaften eindeutig bewerten zu können, müssen sie in verschiedenen Umwelten kultiviert und

bewertet werden. Ist jedoch ein (oder mehrere) Marker für eine Eigenschaft bekannt, so kann eine einzelne Pflanze aufgrund ihres Genotyps ausgewählt werden.

Des Weiteren ist die gezielte Einkreuzung von neuen Merkmalen mit klassischen Zuchtmethoden nur schwer möglich. Sind jedoch diagnostische Marker für gewünschte Eigenschaften verfügbar, so ist die Markergestützte Selektion eine effiziente Methode um gezielt auf diese Eigenschaften zu züchten. So wird die Markergestützte Selektion bereits erfolgreich in landwirtschaftlichen Kulturen eingesetzt, wie z.B. in *Triticum* sp. zur *Fusarium*-Resistenz (Buerstmayr et al. (2009)), in *Solanum tuberosum* zur *Phytophthora infestans*-Resistenz (Park et al. (2009)), in *Lycopersicon esculentum* zur Virus-Resistenz (Nascimento et al. (2009)) und in *Hordeum vulgare* zur Erhöhung der Malzqualität (Emebiri et al. (2009)).

Aufgrund der langen Reproduktionszeit von *Phalaenopsis* können durch Markergestützte Selektion mehrere Jahre und somit auch Kosten bei der Züchtung neuer Sorten eingespart werden.

Literaturverzeichnis

- Abdurakhmonov, I. Y. and A. Abdukarimov (2008). Application of association mapping to understanding the genetic diversity of plant germplasm resources. *International Journal of Plant Genomics* 2008, 1–19.
- Achleitner, A., N. A. Tinker, E. Zechner, and H. Buerstmayr (2008). Genetic diversity among oat varieties of worldwide origin and associations of AFLP markers with quantitative traits. *Theor Appl Genet* 117(7), 1041–1053.
- Agarwal, M., N. Shrivastava, and H. Padh (2008). Advances in molecular marker techniques and their applications in plant sciences. *Plant Cell Reports* 27(4), 617–631.
- Ajjikuttira, P. A., C. L. Lim-Ho, M. H. Woon, K. H. Ryu, C. A. Chang, C. S. Loh, and S. M. Wong (2002). Genetic variability in the coat protein genes of two orchid viruses: *Cymbidium* mosaic virus and *Odontoglossum* ringspot virus. *Archives of Virology* 147(10), 1943–1954.
- Arditti, J. (1992). *Fundamentals of Orchid Biology*. Number ISBN-13: 978-0471549062. John Wiley and Sons Inc.
- Arends, J. (1970). Cytological observations on genome homology in eight interspecific hybrids of *Phalaenopsis*. *Genetica* 41(1), 88–100.
- Asins, M. (2002). Present and future of quantitative trait locus analysis in plant breeding. *Plant Breeding* 121(4), 281–291.
- Barcaccia, G., S. Meneghetti, E. Albertini, L. Triest, and M. Lucchin (2003). Linkage mapping in tetraploid willows: segregation of molecular markers and estimation of linkage phases support an allotetraploid structure for *Salix alba* x *Salix fragilis* interspecific hybrids. *Heredity* 90(2), 169–180.
- Baudry, E., C. Kerdelhue, H. Innan, and W. Stephan (2001). Species and recombination effects on DNA variability in the tomato genus. *Genetics* 158(4), 1725–1735.

- Becher, S., K. Steinmetz, K. Weising, S. Boury, D. Peltier, J. Renou, G. Kahl, and K. Wolff (2000). Microsatellites for cultivar identification in *Pelargonium*. *Theoretical and Applied Genetics* 101(4), 643–651.
- Belaj, A., Z. Satovic, G. Cipriani, L. Baldoni, R. Testolin, L. Rallo, and I. Trujillo (2003). Comparative study of the discriminating capacity of RAPD, AFLP and SSR markers and of their effectiveness in establishing genetic relationships in olive. *Theoretical and Applied Genetics* 107(4), 736–744.
- Bennett, M. D. (2005). Nuclear DNA amounts in angiosperms: progress, problems and prospects. *Annals of Botany* 95(1), 45–90.
- Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Research* 27(2), 573–580.
- Bernardo, R. (2008). Molecular markers and selection for complex traits in plants: learning from the last 20 years. *Crop Science* 48(5), 1649–1664.
- Besenbacher, S., T. Mailund, and M. H. Schierup (2008). Local phylogeny mapping of quantitative traits: higher accuracy and better ranking than single-marker association in genomewide scans. *Genetics* 181(2), 747–753.
- Blair, M., V. Hedetale, and S. McCouch (2002). Fluorescent-labeled microsatellite panels useful for detecting allelic diversity in cultivated rice (*Oryza sativa* L.). *Theoretical and Applied Genetics* 105(2-3), 449–457.
- Botstein, D., R. White, M. Skolnick, and R. Davis (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American Journal of Human Genetics* 32(3), 314.
- Bowman, J. (1997). Evolutionary conservation of angiosperm flower development at the molecular and genetic levels. *Journal of Biosciences* 22(4), 515–527.
- Bradbury, P. J., Z. Zhang, D. E. Kroon, T. M. Casstevens, Y. Ramdoss, and E. S. Buckler (2006). TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* 23(19), 2633–2635.
- Breseghello, F. (2005). Association mapping of kernel size and milling quality in wheat (*Triticum aestivum* L.) cultivars. *Genetics* 172(2), 1165–1177.
- Brondani, R., C. Brondani, R. Tarchini, and D. Grattapaglia (1998). Development, characterization and mapping of microsatellite markers in *Eucalyptus grandis* and *E. urophylla*. *TAG Theoretical and Applied Genetics* 97(5), 816–827.

- Brown, P. J., W. L. Rooney, C. Franks, and S. Kresovich (2008). Efficient mapping of plant height quantitative trait loci in a *Sorghum* association population with introgressed dwarfing genes. *Genetics* 180(1), 629–637.
- Buckler, E. and J. Thornsberry (2002). Plant molecular diversity and applications to genomics. *Current Opinion in Plant Biology* 5(2), 107–111.
- Buerstmayr, H., T. Ban, and J. Anderson (2009). QTL mapping and marker-assisted selection for *Fusarium* head blight resistance in wheat: a review. *Plant Breeding* 128, 1–26.
- Campbell, V., G. Rowe, T. BEEBEE, and M. HUTCHINGS (2007). Genetic differentiation amongst fragrant orchids (*Gymnadenia conopsea* SL) in the british isles. *Botanical Journal of the Linnean Society* 155(3), 349–360.
- Cavanagh, C., M. Morell, I. Mackay, and W. Powell (2008). From mutations to MAGIC: resources for gene discovery, validation and delivery in crop plants. *Current Opinion in Plant Biology* 11(2), 215–221.
- Chai, D., S. M. Lee, J. H. Ng, and H. Yu (2007). L-Methionine sulfoximine as a novel selection agent for genetic transformation of orchids. *Journal of Biotechnology* 131(4), 466–472.
- Chakravarti, A., L. Lasher, and J. Reefer (1991). A maximum likelihood method for estimating genome length using genetic linkage data. *Genetics* 128(1), 175–182.
- Chalker-Scott, L. (1999). Environmental significance of anthocyanins in plant stress responses. *Photochemistry and Photobiology* 70(1), 1–9.
- Chang, Y.-K., R. E. Veilleux, and M. J. Iqbal (2009). Analysis of genetic variability among *Phalaenopsis* species and hybrids using Amplified Fragment Length Polymorphism. *Journal of the American Society for Horticultural Science* 134(1), 58–66.
- Chen, C., Q. Yu, S. Hou, Y. Li, M. Eustice, R. Skelton, O. Veatch, R. Herdes, L. Diebold, and J. Saw (2007). Construction of a sequence-tagged high-density genetic map of papaya for comparative structural and evolutionary genomics in brassicales. *Genetics* 177(4), 2481.
- Chen, D., B. Guo, S. Hexige, T. Zhang, D. Shen, and F. Ming (2007). SQUA-like genes in the orchid *Phalaenopsis* are expressed in both vegetative and reproductive tissues. *Planta* 226(2), 369–380.

- Christenson, E. A. (2001). *Phalaenopsis: a monograph*. Number ISBN-13: 978-0881924947. Timber Press Inc.
- Churchill, G. and R. Doerge (1994). Empirical threshold values for quantitative trait mapping. *Genetics* 138, 963–971.
- Cockram, J., J. White, F. J. Leigh, V. J. Lea, E. Chiapparino, D. A. Laurie, I. J. Mackay, W. Powell, and D. M. O’sullivan (2008). Association mapping of partitioning loci in barley. *BMC Genetics* 9(1), 16.
- Collard, B. and D. Mackill (2008). Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. *Euphytica* 363(1491), 557.
- Collard, B. C. Y., M. Z. Z. Jahufer, J. B. Brouwer, and E. C. K. Pang (2005). An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. *Euphytica* 142(1-2), 169–196.
- Corder, E., A. Saunders, N. Risch, and W. Strittmatter (1994). Protective effect of apolipoprotein E type 2 allele for late onset Alzheimer disease. *Nature Genetics* 7, 180–184.
- Devlin, B., K. Roeder, and L. Wasserman (2000). Genomic control for association studies: a semiparametric test to detect excess-haplotype sharing. *Biostatistics* 1(4), 369–387.
- D’hoop, B. B., M. J. Paulo, R. A. Mank, H. J. Eck, and F. A. Eeuwijk (2008). Association mapping of quality traits in potato (*Solanum tuberosum* L.). *Euphytica* 161(1-2), 47–60.
- Dice, L. (1945). Measures of the amount of ecologic association between species. *Ecology* 26(3), 297–302.
- Dixon, R., M. Harrison, and C. Lamb (1994). Early events in the activation of plant defense responses. *Annual Review of Phytopathology* 32(1), 479–501.
- Doyle, J. J. and J. L. Doyle (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* 19, 11–15.
- Dugo, M. L., Z. Satovic, T. Millán, J. I. Cubero, D. Rubiales, A. Cabrera, and A. M. Torres (2005). Genetic mapping of QTLs controlling horticultural traits in diploid roses. *Theoretical and Applied Genetics* 111(3), 511–520.

- Emebiri, L., P. Michael, D. B. Moody, F. C. Ogbonnaya, and C. Black (2009). Pyramiding QTLs to improve malting quality in barley: gains in phenotype and genetic diversity. *Mol Breeding* 23(2), 219–228.
- Eun, A., M. Seoh, and S. Wong (2000). Simultaneous quantitation of two orchid viruses by the TaqMan® real-time RT-PCR. *Journal of Virological Methods* 87(1-2), 151–160.
- Evanno, G., S. Regnaut, and J. Goudet (2005). Detecting the number of clusters of individuals using the software *structure*: a simulation study. *Molecular Ecology* 14(8), 2611–2620.
- Falush, D., M. Stephens, and J. Pritchard (2007). Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Molecular Ecology Notes* 7, 574–578.
- Felsenstein, J. (1985). Confidence-limits on phylogenies: an approach using the bootstrap. *Evolution* 39(4), 783–791.
- Felsenstein, J. (2005). Phylip phylogeny inference package version 3.6. *Department of Genome Sciences, University of Washington, Seattle*.
- Ferreira, A., M. Silva, L. Silva, and C. Cruz (2006). Estimating the effects of population size and type on the accuracy of genetic maps. *Genetics and Molecular Biology* 29, 187–192.
- Flint-Garcia, S. A., J. M. Thornsberry, E. S, and B. Iv (2003). Structure of linkage disequilibrium in plants. *Annu. Rev. Plant. Biol.* 54(1), 357–374.
- Garcia-Mas, J., M. Oliver, H. Gomez-Paniagua, and M. D. Vicente (2000). Comparing AFLP, RAPD and RFLP markers for measuring genetic diversity in melon. *Theoretical and Applied Genetics* 101(5), 860–864.
- Gawenda, I. and T. Debener (2009). Genetic diversity of *Osteospermum* genotypes analysed by AFLP and chloroplast SSR markers. *European Journal of Horticultural Science* 74(2), 86–94.
- Gebhardt, C., A. Ballvora, B. Walkemeier, P. Oberhagemann, and K. Schöler (2004). Assessing genetic potential in germplasm collections of crop plants by marker-trait association: a case study for potatoes with quantitative variation of resistance to late blight and maturity type. *Molecular Breeding* 13(1), 93–102.

- Gow, W.-P., J.-T. Chen, and W.-C. Chang (2008). Influence of growth regulators on direct embryo formation from leaf explants of *Phalaenopsis* orchids. *Acta Physiol Plant* 30(4), 507–512.
- Griesbach, R. (1985). Polyploidy in orchid improvement. *The Journal of Heredity* 76, 74–75.
- Griesbach, R. (2002). Development of *Phalaenopsis* orchids for the mass-market. *Trends in new crops and new uses. ASHS Press, Alexandria, VA*, 458–465.
- Gruss, O. and M. Wolff (1995). *Phalaenopsis*. Number ISBN-13: 978-3800165513. Ulmer.
- Guo, B., S. Hexige, T. Zhang, J. K. Pittman, D. Chen, and F. Ming (2007). Cloning and characterization of a PI-like MADS-Box gene in *Phaldenopsis* orchid. *Journal of Biochemistry and Molecular Biology* 40(6), 845–852.
- Guo-Qing, G., H. Guo-Hao, and L. Yang-Rui (2003). Microsatellite enrichment from AFLP fragments by magnetic beads. *Acta Botanica Sinica* 45(11), 1266–1269.
- Gupta, P., H. Balyan, P. Sharma, and B. Ramesh (1996). Microsatellites in plants: a new class of molecular markers. *Current Science (Bangalore)* 70(1), 45–54.
- Gupta, P. K., S. Rustgi, and P. L. Kulwal (2005). Linkage disequilibrium and association studies in higher plants: Present status and future prospects. *Plant Molecular Biology* 57(4), 461–485.
- Hackett, C., J. Bradshaw, R. Meyer, J. McNicol, D. Milbourne, and R. Waugh (1998). Linkage analysis in tetraploid species: a simulation study. *Genetics Research* 71(02), 143–153.
- Han, T., H. V. Eck, M. D. Jeu, and E. Jacobsen (1999). Optimization of AFLP fingerprinting of organisms with a large-sized genome: a study on *Alstroemeria* spp. *Theoretical and Applied Genetics* 98(3), 465–471.
- Hansen, M., T. Kraft, S. Ganestam, T. Saell, and N. Nilsson (2001). Linkage disequilibrium mapping of the bolting gene in sea beet using AFLP markers. *Genetics Research* 77(01), 61–66.
- Hardy, G. (1908). Mendelian proportions in a mixed population. *Science* 28(706), 49–50.
- Heinrichs, F. (2008). AIPH / Union Fleurs:International Statistics Flowers and Plants. *Inst. f. Entw.- und Agrarökonomik* (56), 43–44.

- Hill, W. and B. Weir (1994). Maximum-likelihood estimation of gene location by linkage disequilibrium. *American Journal of Human Genetics* 54, 705–714.
- Holsinger, K., P. Lewis, and D. Dey (2002). A Bayesian approach to inferring population structure from dominant markers. *Molecular Ecology* 11(7), 1157–1164.
- Hübner, S., M. Höffken, E. Oren, G. Haseneyer, N. Stein, A. Graner, K. Schmid, and E. Fridman (2009). Strong correlation of wild barley (*Hordeum spontaneum*) population structure with temperature and precipitation variation. *Molecular Ecology* 18(7), 1523–1536.
- Jaccard, P. (1908). Nouvelles recherches sur la distribution florale. *Bull. Soc. Vaud. Sci. Nat.* 44, 223–270.
- Jorde, L. (2000). Linkage disequilibrium and the search for complex disease genes. *Genome Research* 10(10), 1435–1444.
- Julier, B., S. Flajoulot, P. Barre, G. Cardinet, S. Santoni, T. Huguet, and C. Huyghe (2003). Construction of two genetic linkage maps in cultivated tetraploid alfalfa (*Medicago sativa*) using microsatellite and AFLP markers. *BMC Plant Biology* 3(9).
- Kim, M.-S., C. W. Morden, Y. Sagawa, and J.-Y. Kim (2003). The phylogeny of *Phalaenopsis* species. *Proceedings of NIOC*, 1–3.
- Kosambi, D. (1944). The estimation of map distances from recombination values. *Annual Eugenics*, 172–175.
- Kraakman, A., R. Niks, P. V. den Berg, and P. Stam (2004). Linkage disequilibrium mapping of yield and yield stability in modern spring barley cultivars. *Genetics* 168, 435–446.
- Kumapatla, S., M. Manley, E. Horne, M. Gupta, and S. Thompson (2004). An improved enrichment procedure to develop multiple repeat classes of cotton microsatellite markers. *Plant Molecular Biology Reporter* 22(1), 85.
- Latch, E. K., G. Dharmarajan, J. C. Glaubitz, and O. E. Rhodes (2006). Relative performance of bayesian clustering software for inferring population substructure and individual assignment at low levels of population differentiation. *Conservation Genetics* 7(2), 295–302.

- Laurentin, H. (2009). Data analysis for molecular characterization of plant genetic resources. *Genet Resour Crop Evol* 56(2), 277–292.
- Lerceteau-Köhler, E., G. Guerin, F. Laigret, and B. Denoyes-Rothan (2003). Characterization of mixed disomic and polysomic inheritance in the octoploid strawberry (*Fragaria x ananassa*) using AFLP mapping. *Theoretical and Applied Genetics* 107(4), 619–628.
- Lewontin, R. (1964). The interaction of selection and linkage. *Genetics* 49, 49–67.
- Liao, L., I. Pan, Y. Chan, Y. Hsu, W. Chen, and M. Chan (2004). Transgene silencing in *Phalaenopsis* expressing the coat protein of *Cymbidium* Mosaic Virus is a manifestation of RNA-mediated resistance. *Molecular Breeding* 13(3), 229–242.
- Lin, S., L. Hsiao-Ching, W.-H. Chen, C.-C. Chen, Y. Kao, Y.-M. Fu, Y.-H. Chen, and T. Lin (2001). Nuclear DNA contents of *Phalaenopsis* sp. and *Doritis pulcherrima*. *Journal of the American Society for Horticultural Science* 126(2), 195–199.
- Lin, T.-C. (2002). Isolation and characterization of microsatellites in *Phalaenopsis* orchids. Master’s thesis, National Cheng Kung University.
- Linde, M., A. Hattendorf, H. Kaufmann, and T. Debener (2006). Powdery mildew resistance in roses: QTL mapping in different environments using selective genotyping. *Theoretical and Applied Genetics* 113(6), 1081–1092.
- Liu, B. (1998). *Statistical genomics: linkage, mapping, and QTL analysis*. Number ISBN 0-8493-3166-8. CRC Press LLC.
- Ma, H., M. Pooler, and R. Griesbach (2009, Jan). Anthocyanin regulatory/structural gene expression in phalaenopsis. *Journal of the American Society for Horticultural Science* 134(1), 88–96.
- Maccaferri, M., M. Sanguineti, E. Noli, and R. Tuberosa (2005). Population structure and long-range linkage disequilibrium in a durum wheat elite collection. *Molecular Breeding* 15, 271–289.
- Mackay, I. and W. Powell (2007). Methods for linkage disequilibrium mapping in crops. *Trends in Plant Science* 12(2), 57–63.
- Meudt, H. and A. Clarke (2007). Almost forgotten or latest practice? AFLP applications, analyses and advances. *Trends in Plant Science* 12(3), 106–117.

- Miller, M. and B. Yuan (1997). Semiautomated resolution of overlapping stutter patterns in genomic microsatellite analysis. *Analytical Biochemistry* 251(1), 50–56.
- Moises, H., L. Yang, H. Kristbjarnarson, C. Wiese, W. Byerley, F. Macciardi, V. Arolt, D. Blackwood, X. Liu, and B. Sjögren (1995). An international two-stage genome-wide search for schizophrenia susceptibility genes. *Nature Genetics* 11(3), 321–324.
- Moose, S. P. and R. H. Mumm (2008). Molecular plant breeding as the foundation for 21st century crop improvement. *Plant Physiology* 147(3), 969–977.
- Mueller, U. and L. Wolfenbarger (1999). AFLP genotyping and fingerprinting. *Trends in Ecology & Evolution* 14(10), 389–394.
- Nascimento, I., W. Maluf, and A. Figueira (2009). Marker assisted identification of tospovirus resistant tomato genotypes in segregating progenies. *Scientia Agrícola* 66(3), 298–303.
- Nei, M. and W. Li (1979). Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences* 76(10), 5269–5273.
- Niehues, R. (2009). Die wichtigsten Pflanzen im grünen Einzelhandel. *DEGA* (4), 12–13.
- Nordborg, M. and P. Donnelly (1997). The coalescent process with selfing. *Genetics* 146, 1185–1195.
- Nordborg, M. and H. Innan (2002). Molecular population genetics. *Current Opinion in Plant Biology* 5(1), 69–73.
- Nordborg, M. and D. Weigel (2008). Next-generation genetics in plants. *Nature* 456(7223), 720–723.
- Oraguzie, N. C., E. H. Rikkerink, S. E. Gardiner, and H. N. D. Silva (Eds.) (2006). *Association Mapping in Plants*. Number ISBN: 0-387-35844-7. Springer Science and Business Media, LLC.
- Padolina, J. M. (2006). Phylogenetic reconstruction of *Phalaenopsis* using nuclear and chloroplast DNA sequence data. Ph. D. thesis, University of Texas.
- Page, R. D. M. (1996). TREEVIEW: An application to display phylogenetic trees on personal computers. *Computer Applications in the Biosciences* 12, 357–358.

- Park, T.-H., V. G. A. A. Vleeshouwers, E. Jacobsen, E. V. D. Vossen, and R. G. F. Visser (2009). Molecular breeding for resistance to *Phytophthora infestans* (Mont.) de Bary in potato (*Solanum tuberosum*): a perspective of cisgenesis. *Plant Breeding* 128(2), 109–117.
- Powell, W., M. Morgante, C. Andre, M. Hanafey, J. Vogel, S. Tingey, and A. Rafalski (1996). The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. *Molecular Breeding* 2(3), 225–238.
- Presterl, T., N. Stein, D. Perovic, F. Ordon, and A. Graner (2007). Effects of Introgression and recombination on haplotype structure and linkage disequilibrium surrounding a locus encoding Bymovirus Resistance in Barley. *Genetics* 175, 805–817.
- Price, A. L., N. J. Patterson, R. M. Plenge, M. E. Weinblatt, N. A. Shadick, and D. Reich (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics* 38(8), 904–909.
- Pritchard, J., M. Stephens, and P. Donnelly (2000). Inference of Population Structure Using Multilocus Genotype Data. *Genetics* 155(2), 945–959.
- Pritchard, J., M. Stephens, N. Rosenberg, and P. Donnelly (2000). Association Mapping in Structured Populations. *The American Journal of Human Genetics* 67(1), 170–181.
- Quarrie, S. A., A. Steed, C. Calestani, A. Semikhodskii, C. Lebreton, C. Chinoy, N. Steele, D. Pljevljakusić, E. Waterman, J. Weyen, J. Schondelmaier, D. Z. Habash, P. Farmer, L. Saker, D. T. Clarkson, A. Abugalieva, M. Yessimbekova, Y. Turuspekov, S. Abugalieva, R. Tuberosa, M.-C. Sanguineti, P. A. Hollington, R. Aragués, A. Royo, and D. Dodig (2005). A high-density genetic map of hexaploid wheat (*Triticum aestivum* L.) from the cross Chinese Spring SQ1 and its use to compare QTLs for grain yield across a range of environments. *Theoretical and Applied Genetics* 110(5), 865–880.
- Rafalski, A. and M. Morgante (2004). Corn and humans: recombination and linkage disequilibrium in two genomes of similar size. *Trends in Genetics* 20(2), 103–111.
- Rao, N. (2004). Plant genetic resources: Advancing conservation and use through biotechnology. *African Journal of Biotechnology* 3(2), 136–145.
- Reif, J. C., S. Hamrit, M. Heckenberger, W. Schipprack, H. P. Maurer, M. Bohn, and A. E. Melchinger (2005). Genetic structure and diversity of European flint maize

- populations determined with SSR analyses of individuals and bulks. *Theoretical and Applied Genetics* 111(5), 906–913.
- Remington, D., J. Thornsberry, Y. Matsuoka, L. Wilson, S. Whitt, J. Doebley, S. Kresovich, M. Goodman, and E. Buckler (2001). Structure of linkage disequilibrium and phenotypic associations in the maize genome. *Proceedings of the National Academy of Sciences* 98(20), 11479.
- Remington, D., R. Whetten, B. Liu, and D. O’malley (1999). Construction of an AFLP genetic map with nearly complete genome coverage in *Pinus taeda*. *Theoretical and Applied Genetics* 98(8), 1279–1292.
- Rodrigues, K. F. and S. V. Kumar (2008). Isolation and characterization of microsatellite loci in *Phalaenopsis gigantea*. *Conservation Genetics* 10(3), 559–562.
- Rozen, S. and H. J. Skaletsky (2000). Primer3 on the www for general users and for biologist programmers. In: Kravetz S, Misener S (eds), *Bioinformatic Methods and Protocols: Methods in Molecular Biology*, Humana Press, 365–386.
- Saitou, N. and M. Nei (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* 4(4), 406–425.
- Salvi, S., G. Sponza, M. Morgante, D. Tomes, X. Niu, K. Fengler, R. Meeley, E. Ananiev, S. Svitashhev, and E. Bruggemann (2007). Conserved noncoding genomic sequences associated with a flowering-time quantitative trait locus in maize. *Proceedings of the National Academy of Sciences* 104(27), 11376–11381.
- Schuelke, M. (2000). An economic method for the fluorescent labeling of PCR fragments. *Nature Biotechnology* 18(2), 233–234.
- Shete, S., H. Tiwari, and R. Elston (2000). On Estimating the Heterozygosity and Polymorphism Information Content Value. *Theoretical Population Biology* 57(3), 265–271.
- Shrestha, B. R., K. Tokuhara, and M. Mii (2007). Plant regeneration from cell suspension-derived protoplasts of *Phalaenopsis*. *Plant Cell Reports* 26(6), 719–725.
- Simko, I. (2004). One potato, two potato: haplotype association mapping in autotetraploids. *Trends in Plant Science* 9(9), 441–448.
- Skøt, L., M. O. Humphreys, I. Armstead, S. Heywood, K. P. Skøt, R. Sanderson, I. D. Thomas, K. H. Chorlton, and N. R. S. Hamilton (2005). An association mapping

- approach to identify flowering time genes in natural populations of *Lolium perenne* (L.). *Molecular Breeding* 15(3), 233–245.
- Smith, J., E. Chin, H. Shu, O. Smith, S. Wall, M. Senior, S. Mitchell, S. Kresovich, and J. Ziegler (1997). An evaluation of the utility of SSR loci as molecular markers in maize (*Zea mays* L.): comparisons with data from RFLPs and pedigree. *Theoretical and Applied Genetics* 95(1), 163–173.
- Sorkheh, I. K., L. Malysheva-Otto, M. Wirthensohn, S. Tarkesh-Esfahani, and P. Martínez-Gómez (2008). Linkage disequilibrium, genetic association mapping and gene localization in crop plants. *Genetics and Molecular Biology* 31(4).
- Squirrell, J., P. M. Hollingsworth, M. Woodhead, J. Russell, A. J. Lowe, M. Gibby, and W. Powell (2003). How much effort is required to isolate nuclear microsatellites from plants? *Molecular Ecology* 12(6), 1339–1348.
- Stam, P. (1993). Construction of integrated genetic linkage maps by means of a new computer package: JoinMap. *Plant Journal* 3(5), 739–744.
- Stein, J., S. C. Pessino, E. J. Martínez, M. P. Rodriguez, L. A. Siena, C. L. Quarín, and J. P. A. Ortiz (2007). A genetic map of tetraploid *Paspalum notatum* Flugge (bahiagrass) based on single-dose molecular markers. *Molecular Breeding* 20(2), 153–166.
- Stich, B., A. E. Melchinger, M. Heckenberger, J. Möhring, A. Schechert, and H.-P. Piepho (2008). Association mapping in multiple segregating populations of sugar beet (*Beta vulgaris* L.). *Theoretical and Applied Genetics* 117(7), 1167–1179.
- Stich, B., H.-P. Piepho, B. Schulz, and A. E. Melchinger (2008). Multi-trait association mapping in sugar beet (*Beta vulgaris* L.). *Theoretical and Applied Genetics* 117(6), 947–954.
- Sweet, H. (1980). *The genus Phalaenopsis*. Day Printing Corp, California.
- Tanksley, S. (1993). Mapping polygenes. *Annual Review of Genetics* 27(1), 205–233.
- Team, R. D. C. (2008). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*.
- Tsai, C., S. Huang, P. Huang, and C. Chou (2003). Phylogeny of the genus *Phalaenopsis* (Orchidaceae) with emphasis on the subgenus *Phalaenopsis* based on the sequences of the internal transcribed spacers 1 and 2 of rDNA. *Journal of Horticultural Science and Biotechnology* 78(6), 879–887.

- Tsai, C. C., S. C. Huang, and C. H. Chou (2005). Molecular phylogeny of *Phalaenopsis* Blume (Orchidaceae) based on the internal transcribed spacer of the nuclear ribosomal DNA. *Plant Syst. Evol.* 256(1-4), 1–16.
- Tsai, W.-C. (2005). PeMADS6, a GLOBOSA/PISTILLATA-like gene in *Phalaenopsis equestris* involved in petaloid formation, and correlated with flower longevity and ovary development. *Plant and Cell Physiology* 46(7), 1125–1139.
- Tsai, W.-C., Y.-Y. Hsiao, Z.-J. Pan, C.-C. Hsu, Y.-P. Yang, W.-H. Chen, and H.-H. Chen (2008). Molecular biology of orchid flowers: With emphasis on *Phalaenopsis*. *Adv Bot Res* 47, 99–145.
- VanOoijen, J. (2006). *JoinMap® 4, Software for the calculation of genetic linkage maps in experimental populations*. Wageningen, Netherlands: Kyazma B.V.
- VanOoijen, J. (2009). *MapQTL 6, Software for mapping of quantitative trait loci in experimental populations of diploid species*. Wageningen, Netherlands: Kyazma B.V.
- Verma, P., D. Chakrabarty, S. Jena, D. Mishra, P. Singh, S. Sawant, and R. Tuli (2009). The extent of genetic diversity among *Vanilla* species: Comparative results for RAPD and ISSR. *Industrial Crops & Products* 29(2-3), 581–589.
- Voorrips, R. (2002). MapChart: software for the graphical presentation of linkage maps and QTLs. *Journal of Heredity* 93(1), 77–78.
- Vos, P., R. Hogers, M. Bleeker, M. Reijans, T. V. D. Lee, M. Hornes, A. Frijters, J. Pot, J. Peleman, and M. Kuiper (1995). AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research* 23(21), 4407–4414.
- Vuylsteke, M., J. D. Peleman, and M. J. V. Eijk (2007). AFLP technology for DNA fingerprinting. *Nature Protocols* 2(6), 1387–1398.
- Wang, H.-Z., Z.-X. Wu, J.-J. Lu, N.-N. Shi, Y. Zhao, Z.-T. Zhang, and J.-J. Liu (2009). Molecular diversity and relationships among *Cymbidium goeringii* cultivars based on inter-simple sequence repeat (ISSR) markers. *Genetica* 136(3), 391–399.
- Waples, R. S. and O. Gaggiotti (2006). What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. *Molecular Ecology* 15(6), 1419–1439.

- Waugh, R., J. Jannink, G. Muehlbauer, and L. Ramsay (2009). The emergence of whole genome association scans in barley. *Current Opinion in Plant Biology* 12, 218–222.
- Weinberg, W. (1909). Über Vererbungsgesetze beim Menschen. *Molecular and General Genetics MGG* 2(1), 276–330.
- Weising, K., R. Fung, D. Keeling, R. Atkinson, and R. Gardner (1996). Characterisation of microsatellites from *Actinidia chinensis*. *Molecular Breeding* 2(2), 117–131.
- Weising, K., H. Nybom, K. Wolff, and G. Kahl (2005). *DNA fingerprinting in plants: principles, methods, and applications* (2nd ed.). CRC Press.
- Wills, D. M. and J. M. Burke (2007). Quantitative trait locus analysis of the early domestication of sunflower. *Genetics* 176(4), 2589–2599.
- Winkelmann, T., T. Geier, and W. Preil (2006). Commercial in vitro plant production in Germany in 1985–2004. *Plant Cell Tiss Organ Cult* 86(3), 319–327.
- Wu, K., W. Burnquist, M. Sorrells, T. Tew, P. Moore, and S. Tanksley (1992). The detection and estimation of linkage in polyploids using single-dose restriction fragments. *Theoretical and Applied Genetics* 83(3), 294–300.
- Xiang, N., Y. Hong, and L. Lam-Chan (2003). Genetic analysis of tropical orchid hybrids (*Dendrobium*) with fluorescence amplified fragment-length polymorphism (AFLP). *Journal of the American Society for Horticultural Science* 128(5), 731–735.
- Yan, Z., P. B. Visser, T. Hendriks, T. W. Prins, P. Stam, and O. Dolstra (2007). QTL analysis of variation for vigour in rose. *Euphytica* 154(1-2), 53–62.
- Yang, R., N. Yi, and S. Xu (2006). Box–Cox transformation for QTL mapping. *Genetica* 128(1-3), 133–143.
- Yu, J., G. Pressoir, W. H. Briggs, I. V. Bi, M. Yamasaki, J. F. Doebley, M. D. McMullen, B. S. Gaut, D. M. Nielsen, J. B. Holland, S. Kresovich, and E. S. Buckler (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature Genetics* 38(2), 203–208.
- Yue, G. H., L. T. Lam-Chan, and Y. Hong (2006). Development of simple sequence repeat (SSR) markers and their use in identification of *Dendrobium* varieties. *Molecular Ecology Notes* 6(3), 832–834.

- Zane, L., L. Bargelloni, and T. Patarnello (2002). Strategies for microsatellite isolation: a review. *Mol Ecol* 11(1), 1–16.
- Zeid, M., C.-C. Schön, and W. Link (2003). Genetic diversity in recent elite faba bean lines using AFLP markers. *Theoretical and Applied Genetics* 107(7), 1304–1314.
- Zhu, C., M. Gore, E. S. Buckler, and J. Yu (2008). Status and Prospects of Association Mapping in Plants. *The Plant Genome* 1(1), 5–20.

5 Anhang

5.1 Zusätzlich verwendetes Material

5.1.1 Bakterienstamm

Tabelle 5.1: In der vorliegenden Arbeit verwendete Bakterienstamm

Stamm	Genotyp	Quelle	Verwendung
<i>E. coli</i> DH10B	F- mcrA D(mrr-hsd RMS-mcrBC) f80dlacZDM15 DlacX74 deoR recA1 endA1 araD139 D(ara, leu)7697 galU galK l-rpsL nupG	Invitrogen	Klonierung von PCR Produkten

5.1.2 Verwendete Geräte

Tabelle 5.2: Verwendete Geräte

Gerät	Firma	Bezeichnung
Doppellasersequenzsystem	Licor	DNA Analyzer 4200 und 4300
Elektroporationsgerät	BioRad	Micropulser
Elektroporationsküvetten	Invitrogen	E-cuvets 0.1cm
Fotometer	LAT, Safas	
Geldokumentationssystem	Intas	Intas GDS
Gelelektrophoresekammern	Owl, MWG-Biotech, Kisker, Peqlab, Hybaid	
Inkubationsschüttler	Incubec	IH50
Magnetrührer	IKAMAG	Ikamag RET
Minispin	Neolab	C1301
pH-Messgerät	Hanno Instruments	pH211 Microprocessor
Pipetten	Abimed, Eppendorf	Labmate, Multipette plus, Multichannel8
Pipettier Roboter	Eppendorf	Ep-motion 5000
Reinstwasseranlage	Sartorius	arium G11
Replikationsstempel	Sigma	Replica plater
Schüttler	Edmund Bühler	Swip KS-10
Schwingmühle	Retsch	
Thermocycler	Peqlab	Primus 96 advanced Gradient
Vertikale Elektrophoresesysteme	Owl, Biometra	Maxigel
Vortex	VWR international	444-1372
Waagen	Sartorius, Mettler	Universal, AE260 Delta Range
Wärmeschrank	Memmert	
Wasserbad	LAT	W16
Wankeltisch	Biometra	
Zentrifugen	Hettich	Rotina420R, Universal 32R, Mikro200

5.1.3 Hersteller der Geräte und Chemikalien

Tabelle 5.3: Standorte der Hersteller

Hersteller	Standort	Land
Ambion	Cambridgeshire	UK
BioRad	München	D
Biozym Diagnostik	Hessisch Oldendorf	D
Bock Bio Science GmbH	Bremen	D
Eppendorf	Hamburg	D
Hartmann Analytic	Braunschweig	D
Heraeus Sepatech	Osterode	D
Hettich Zentrifugen	Tuttlingen	D
Invitek	Berlin	D
Invitrogen	Karlsruhe	D
LI-COR	Lincoln	USA
Merck	Darmstadt	D
MWG Biotech	Ebersberg	D
Peqlab Biotechnologie	Erlangen	D
Promega	Madison, Wisconsin	USA
Quiagen	Hilden	D
Retsch	Haan	D
Roche Diagnostics	Mannheim	D
Roth	Karlsruhe	D
Sigma-Aldrich	St. Louis, Missouri	USA
Nerbe plus	Winsen/Luhe	D
Sarstedt	Nürnberg	D
Applied Biosystems	siehe Invitrogen	D
GATC Biotech	Konstanz	D

5.1.4 Zusammensetzung von Medien, Puffer und Lösungen

Tabelle 5.4: Medien

Medium	Zusammensetzung
FB (Freezing Broth)	1 % (w/v) BactoTM Pepton (BD) 0,5 % (w/v) Hefeextrakt 85,5 mM NaCl 36,2 mM K ₂ HPO ₄ 13,2 mM KH ₂ PO ₄ 1,5 mM Natriumcitrat 6,8 mM Ammoniumsulfat 365 M MgSO ₄ 4,4 % (v/v) Glycerin (autoklavieren)
LB (Luria-Bertani) Flüssigmedium	1 % (w/v) BactoTM Pepton (BD) 0,5 % (w/v) Hefeextrakt 171 mM NaCl pH 7,0 mit 5 M NaOH (autoklavieren)
LB (Luria-Bertani) Festmedium	wie Flüssigmedium + 1,5 % (w/v) BactoTM Agar (BD)
SOC	SOB +20 mM Glukose +10 mM MgSO ₄ +10 mM MgCl ₂
SOB	2% (w/v) BactoTM Pepton (BD) 0,5% (w/v) Hefeextrakt 8,5 mM NaCl 2,5 mM KCl pH 7,0 mit 5 M NaOH einstellen (autoklavieren)

Tabelle 5.5: Puffer und Lösungen

Puffer/Lösung	Zusammensetzung
Hybridisierungspuffer	5x SSPE 5x Denhardt's 0,1% SDS 1% Magermilchpulver (frisch hinzu geben)
Denaturierungslösung	0,5mol/l NaOH 1,5mol/l NaCl
Neutralisierungslösung	1,5mol/l NaCl 0,5mol/l Tris (pH 7,4)
TE-Puffer	10 mM TrisHCl pH 8,0 1 mM EDTA pH 8,0
1x TAE-Puffer	40 mM Tris base 1 mM EDTA pH 8,0 mit konzentrierter Essigsäure einstellen
1x TBE	89 mM Tris base 89 mM Borsäure 2 mM EDTA (pH 8,0)
RL-Puffer	100 mM TrisHCl 100 mM Magnesiumacetat 500 mM Kaliumacetat 50 mM DTT Dithiothreitol (77 mg/10 ml) pH 7,5
Williams-Puffer	10 mM Tris pH 8,3 50 mM KCl 2 mM MgCl ₂ 0,001 % Gelatine (BioRad) autoklavieren
TBE-Puffer	89 mM TrisHCl pH 8,0 89 mM Borsäure 2 mM EDTA pH 8,0
CTAB-1 Extraktionspuffer	3% CTAB 100 mM Tris pH 8,0 20 mM EDTA 1,4 M NaCl 1,5% PVP
Denhardt's	1% (w/v) Ficoll 400 1% (w/v) PVP 1% (w/v) Albumin Fraktion V (steril filtrieren)
Ladepuffer PAA-Gele	98% Formamid 10 mM EDTA pH 8,0 0,05% Pararosanilin

Tabelle 5.6: Internetadressen der frei verfügbaren Software

Name	Web Adresse
PHYLIP 3.6.8	http://evolution.genetics.washington.edu/phylip.html
Primer3	http://frodo.wi.mit.edu/primer3/
R 2.8.1	http://www.r-project.org/
STRUCTURE 2.2	http://pritch.bsd.uchicago.edu/structure.html
Tandem Repeats Finder	http://tandem.bu.edu/trf/trf.html
TASSEL 2.1	http://www.maizegenetics.net/
TREEVIEW 1.6.6	http://taxonomy.zoology.gla.ac.uk/rod/treeview.html

5.2 Mikrosatellitenmarker

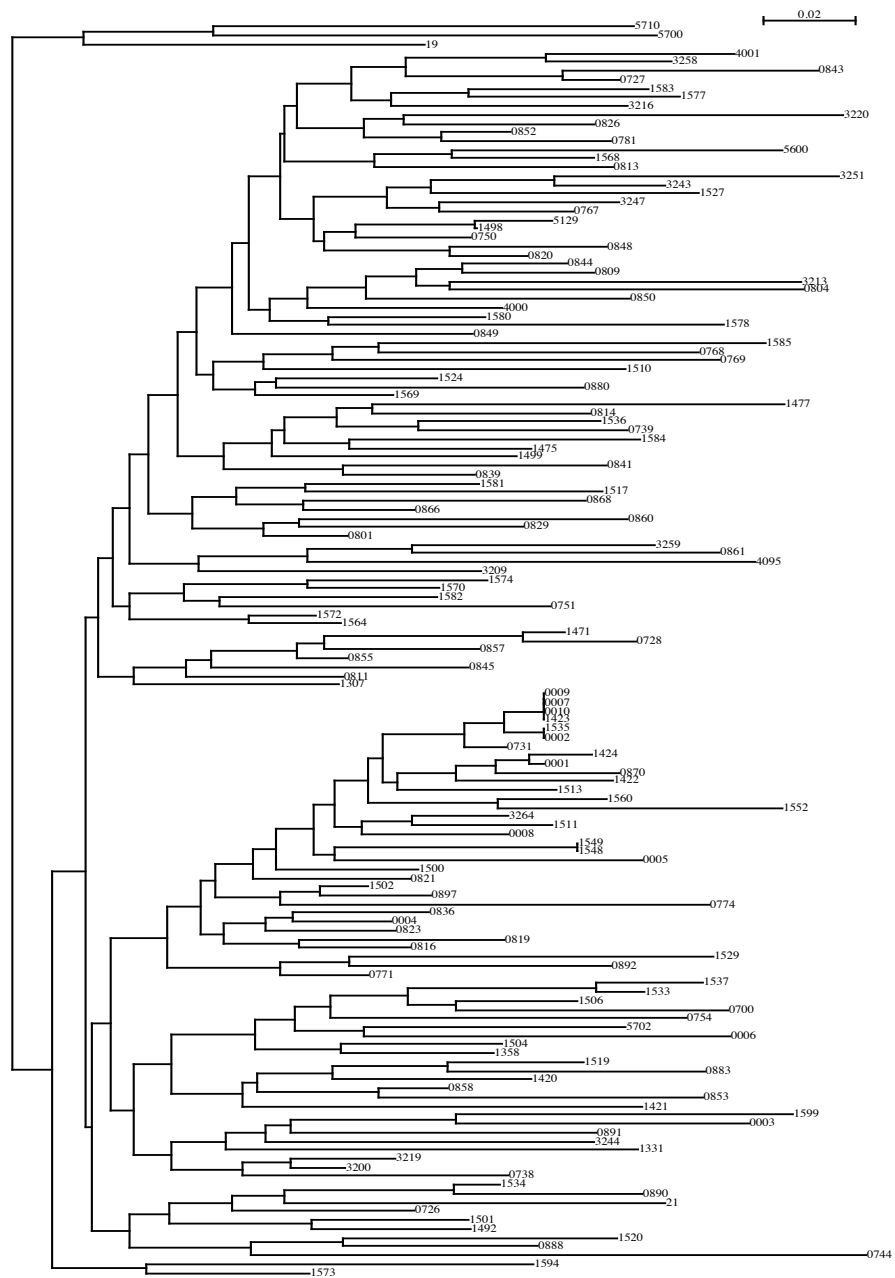


Abbildung 5.1: Neighbour-joining Dendrogram basierend auf sechs Mikrosatellitenmarkern

Tabelle 5.7: Mikrosatellitenprimerpaare entwickelt aus dem Genotyp 0744

Primer bezeichnung	Motiv (perfekt/imperfekt)	Primersequenz	Größe in bp	Annealing temp. in °C
MPH05	(GA)12/15 (imperfekt, unendl. zwei Motive)	l: 5'-CCATGTAGTTCCCCATCACA-3' r: 5'-CCCTCCCAAGCTCCATAATC-3'	190	57°C
MPH10	(GA)11 (imperfekt)	l: 5'-AGAGAGTGGGGACAGGAAT-3' r: 5'-TTCCCTACTTCTTCGCCCTCA-3'	154	62°C
MPH21	(GA)19,5	l: 5'-ACCAGCTTGCAAAGAGTTG-3' r: 5'-TTTCCCTCTCCCTTCACTT-3'	211	58°C
MPH24	(GA)14,5 (imperfekt)	l: 5'-GCTAGAGATCGAAAAAGGTCA-3' r: 5'-TCCCTCTCTCCACTCTTTG-3'	216	60°C
MPH27	(GA)24 (imperfekt)	l: 5'-CGAAGAGGTAAGGGGAAAGG-3' r: 5'-CCTCTATCACACGCTCTGG-3'	216	61°C
MPH39	(GA)10 (imperfekt)	l: 5'-GGAAGGAGTGTTTGGAGGT-3' r: 5'-TCCCTTTGGTCTCACTCGTC-3'	210	59°C
MPH51	(GA)18,5 (imperfekt)	l: 5'-TGATCAGTGGTTGGTGACA-3' r: 5'-GAAGTGGTTGATGGTGGTT-3'	228	59°C
MPH58	(AG)16,5 (imperfekt)	l: 5'-GGGTCAGAGCTTCTTCTCTC-3' r: 5'-TCTTTGTCTCTCTCTTCTCACCA-3'	169	60°C
MPH62	(AG)11 imperfekt	l: 5'-AAGGAGATAAATGATGGAGAGAGAA-3' r: 5'-GCATGGGGACATAAAGAAAGG-3'	244	58°C
MPH64	(AG)28,5 (imperfekt)	l: 5'-CCGGGGGAAAGGATAATG-3' r: 5'-CCTTTCTTTCTTCTCCCTCA-3'	233	57°C
MPH79	(TG)11 (imperfekt)	l: 5'-TGTGCGAGAAGTTGTTGAC-3' r: 5'-CCTGAGTAATGCTTGATCATAGTCTC-3'	242	57°C
MPH88	(T)24 (perfekt)	l: 5'-AGCTTTGCCCTTTGGATTGA-3' r: 5'-AGGCAAGAGACCCCTTTAC-3'	154	56°C
MPH89	(TG)7,5 (perfekt)	l: 5'-TCTTTTCTTGGCACATAATTCA-3' r: 5'-TGGCCCCCAAGAATAGAGAA-3'	250	56°C
MPH95	(TGAA)4 G AG TG (AG)18 (imperfekt, zus.gesetzt)	l: 5'-GCATGAGAGAGGGAGTCATAGA-3' r: 5'-CCCTCACACTCACTCCGACT-3'	228	62°C
MPH120	(AC)7 (AT)4 (AG)8 (zus.gesetzt)	l: 5'-CGACGATCAGACAGTCGAAAG-3' r: 5'-TCCAGAAATGATGGATTCCCTTG-3'	249	57°C
MPH123	(GA)30 (imperfekt)	l: 5'-AACGATGGAGGAGGGAGAGT-3' r: 5'-CCTCTTTGTCATCACTCAACCA-3'	222	58°C
MPH125	(AG)16,5 (perfekt)	l: 5'-AAATAACGGGGGCTAAAA-3' r: 5'-CCCTCTCTCTCCCTACCATC-3'	185	57°C

Tabelle 5.9: veröffentlichte Mikrosatellitenprimerpaare

in der vor- liegenden Arbeit	Primerbezeichnung		Primersequenz	Annealing- temperatur in C	Herkunft
	Originalbezeichnung				
L2	PaAGS002		l: 5'-GCTAGAGATAGAGAGAGAAAGAG-3' r: 5'-TAGCCGACATGCAGAAAG-3'	57°C	Quelle 1
L3	PaAGS003		l: 5'-TCTTTGTGTGTGTGCGTG-3' r: 5'-GTGTCCCGTAGACTTCCCG-3'	58°C	Quelle 1
L4	PaGTS004		l: 5'-GCCCAATTTTCAGTTTTCCTTC-3' r: 5'-TTGCTTCATTGTTTTCCTTCAC-3'	58°C	Quelle 1
L5	PaAGD005		l: 5'-CCATCCCTACACCTGAACAGAG-3' r: 5'-TTCTGCGAGCCCACTACTTC-3'	58°C	Quelle 1
S2	–		l: 5'-AAGGAGATTCAGATGGAGA-3' r: 5'-TTTCAACTTTCTTACCCTAAC-3'	60°C	Quelle 2

Quelle 1: Lin (2002) http://etdncku.lib.ncku.edu.tw/ETD-db/ETD-search/view_etd?URN=etd-0711103-001106
Quelle 2: http://bioinformatics.pcbasc.latrobe.edu.au/cgi-bin/ssr_taxonomy.cgi

5.3 Oligonukleotide für die AFLP-Technik

Tabelle 5.11: Oligonukleotide, die bei der Ligation der DNA-Fragmente als Adaptoren verwendet wurden

Name	Sequenz	Herkunft
<i>Mse</i> I oben	5'-GAC GAT GAG TCC TGA G-3'	MWG Biotech GmbH
<i>Mse</i> I unten	5'-TAC TCA GGA CTC AT-3'	MWG Biotech GmbH
<i>Hind</i> III oben	5'-CTC GTA GAC TGC GTA CC-3'	MWG Biotech GmbH
<i>Hind</i> III unten	5'-AGC TGG TAC GCA GTC TAC-3'	MWG Biotech GmbH

Tabelle 5.12: Grundgerüst der Oligonukleotide, die für die Präamplifikation (P) und für die Endreaktion (E) eingesetzt wurden; N - selektive Nukleotide (siehe Tabelle 5.13)

Name	Verwendung	Sequenz	Herkunft
<i>Mse</i> I+0	P	5'-GAC GAT GAG TCC TGA GTA A-3'	MWG Biotech GmbH
<i>Hind</i> III+0	P	5'-AGA CTG CGT ACC AGC TT-3'	MWG Biotech GmbH
<i>Mse</i> I+NNNN	E	5'-GAT GAG TCC TGA G TAA NNNN-3'	MWG Biotech GmbH
<i>Hind</i> III+NNN ¹	E	5'-GAC TGC GTA CC AGCTT NNN-3'	MWG Biotech GmbH

¹ mit 700 nm oder 800 nm IRD markiert

Tabelle 5.13: AFLP-Primerkombinationen

[illegible]

¹ Schema für die Bezeichnung der Primerkombination: H700AAT → *HindIII*-Primer mit 700er IRDye Markierung und drei selektiven Basen; MAAGC → *MseI*-Primer mit vier selektiven Basen

5.4 Signifikanzschwellen für das Intervallmapping

Gruppe	Anthocyangehalt in der Blüte		Anthocyangehalt im Blatt		Anthocyangehalt in der Lippe		BoxCox Transf. loglip ²		Blütengröße		Rispenlänge	
	mat ³	pat ⁴	mat	pat	mat	pat	mat	pat	mat	pat	mat	pat
1	1,5	4,4	1,4	5,4	1,4	5,9	1,5	5,8	1,6	4,5	1,6	3,6
2	2,8	4	3,1	4,9	4,1	5,8	3,3	5,5	2,7	4,2	2,5	3,3
3	1,4	1,5	1,3	1,3	1,3	1,3	1,3	1,4	1,4	1,4	1,3	1,5
4	2,9	3,5	2,9	4,7	3,9	5	2,8	4,8	2,9	3,7	2,4	2,9
5	3,3	3,8	3,4	4,6	4,7	5,2	3,6	4,5	3,4	3,6	2,8	3,3
6	1,3	3,3	1,3	3,6	1,3	4,6	1,4	3,4	1,4	3,4	1,2	2,8
7	1,5	1,9	1,3	1,8	1,4	1,8	1,4	1,8	1,3	2,1	1,4	1,9
8	1,4	1,6	1,3	1,5	1,2	1,4	1,2	1,4	1,3	1,6	1,3	1,4
9	4	1,6	5	1,5	5,4	1,4	5,2	1,5	4,3	1,5	3,1	1,5
10	1,4	4	1,3	5,1	1,5	5,4	1,4	5,1	1,4	3,9	1,4	3,2
11	3,8	3,4	4	3,9	5	4,8	4,2	4,1	3,8	3,7	2,9	2,6
12	3,8	3,5	4,3	4,2	5,1	4,8	4,9	4,7	3,8	3,8	3	2,7
13	3,7	1,6	4,4	1,6	5,3	1,5	4,8	1,6	4,1	1,8	3	1,6
14	3,4	3,6	4,3	4,7	5	5,2	4,2	4,5	3,7	3,5	2,7	3
15	3,4	3,3	4,3	3,8	4,7	4,7	4,3	3,9	3,9	3,5	2,9	2,7
16	3,5	3,8	4,3	4,5	5	5,3	4,5	5,2	3,8	4,1	2,8	2,9
17	1,9	3,6	1,7	4,5	1,8	5,2	1,7	4,8	1,8	3,8	1,8	2,8
18	3,4	3,1	4,2	3,5	5	4,3	4,8	3,6	4,1	3,4	2,9	2,4
19	3,7	4,1	4,7	5,8	5,3	5,8	5,2	5,7	4	4,1	3	3,6
GW ¹	4,9	5,2	5,9	6,4	6,4	6,8	6,5	6,9	5,1	5,2	4	4,4

Tabelle 5.14: Signifikanzschwellen für das Intervallmapping der QTL-Analyse berechnet mit MapQTL 6; ¹GW - Genomweite Signifikanzschwelle, ²Box-Cox Transformation der Werte 'Anthocyangehalt in der Lippe', ³mat - maternal, ⁴pat - paternal

5.5 Ermittlung der Anzahl der Subpopulationen nach Evanno et al. (2005)

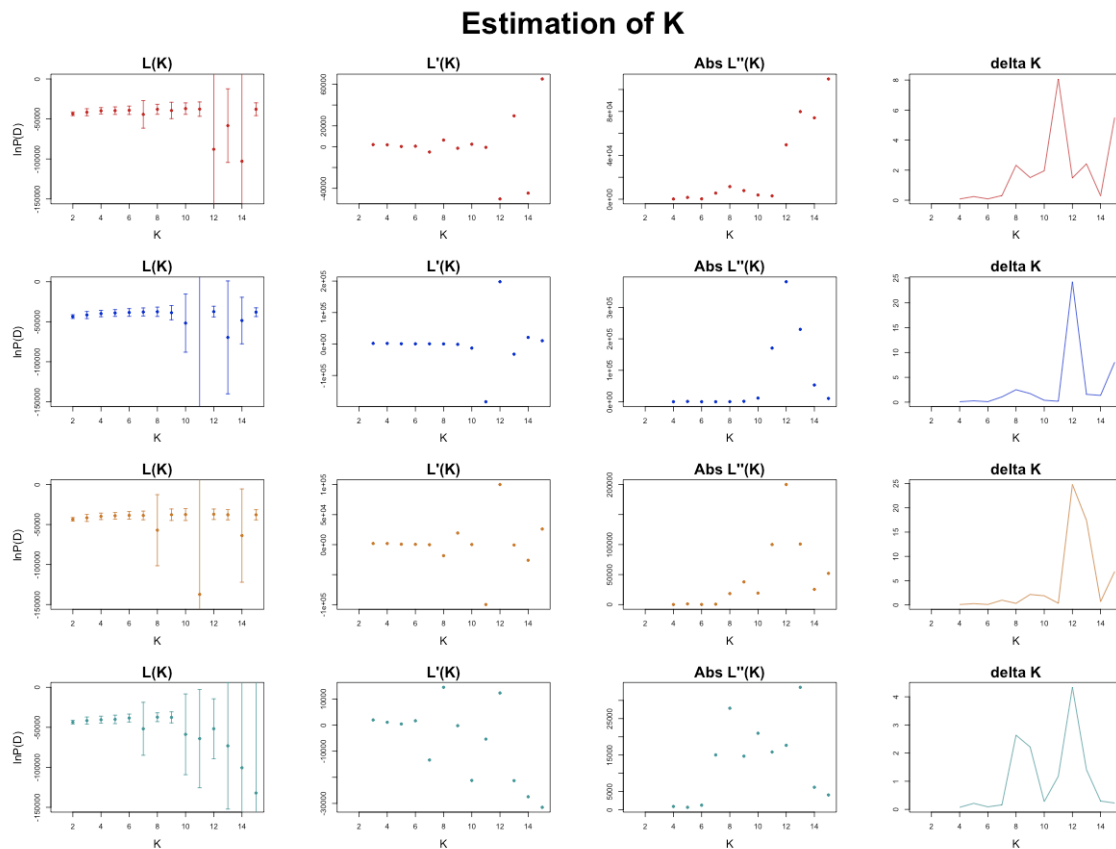


Abbildung 5.2: **graphische Darstellung der Ermittlung der Anzahl der Subpopulationen nach Evanno et al. (2005)**: Die vier unabhängigen Durchläufe aus STRUCTURE 2.2 sind in unterschiedlichen Farben dargestellt. Für jeden Durchlauf wurden die vier Schritte zur Ermittlung der Anzahl an Subpopulationen dargestellt. Die entsprechenden Formeln sind dem Text zu entnehmen. Die wahrscheinlichste Anzahl an Subpopulationen ist jeweils bei dem K zu wählen das einen Peak in der letzten Graphik (ΔK) zeigt. ($L(K)$ entspricht $\ln P(D)$ aus Pritchard et al. (2000))

5.6 R Routine

R Routine zur Berechnung der Marker Eigenschafts Assoziationen

```
str_assoc<-function(fileq,fileg,fileph,phnum){

## Uses likelihood ratio test to measure P(no association).
## Uses Q matrix from STRUCTURE (Pritchard et.al) as a covariate to control
## for historic sub-population structure.
## Calls the function GC to perform simple genomic control on output.
## Returns a matrix of markers (rows) and significances for Structure
## corrected associaiton, raw association and then GC corrected equivalents.
##
## Written by J. White, NIAB, Huntingdon Rd, Cambridge, CB3 0LE.
## Jon.white@niab.com.
##
## Version 1.1.
## 28-01-09
## No guarantee offered or implied.
##
## CAUTION THIS CODE WORKS ON THE ASSUMPTION THAT GENOTYPE IS BINARY.
## IF MULTI_ALLELEIC MARKERS ARE USED THEY WILL NEED TO BE CONVERTED
## TO MULTIPLE BINARY MARKERS.

##read the Q file
q.mat<-read.table(fileq, header=T,row.names=1,quote="")

##Put column names on the structure output.
for (count in 1:ncol(q.mat)){
  colnames(q.mat)[count]<-paste("K",count,sep="")
}
##Attach the q matrix to make it easier to do SA further on.
attach(q.mat)

##Read in the genotype and phenotype.
geno<-read.table(fileg,header=T,row.names=1)
pheno<-read.table(fileph,header=T,row.names=1)
```

```

form<-"
for (qcount in 1:(ncol(q.mat)-1)){
form<-paste(form,"K",qcount,sep="")
form<-paste(form,"+",sep="")
}
## define the size and headers of the output matrix
out<-matrix(NA,ncol=4,nrow=ncol(geno))
rownames(out)<-colnames(geno)[1:ncol(geno)]
colnames (out)<- c("S.Association.P","Raw.Association.P","S.Association.GC.P",
"Raw.Association.GC.P")

for (count in 1:ncol(geno)){

##perform SA
result<-(anova(glm(formula(paste("geno[,",count,"]~",form,"pheno[,phnum]")),
binomial),test="Chisq"))
out[count,1]<- as.numeric(result[ncol(q.mat)+1,5])
result<-anova(glm(geno[,count]~pheno[,phnum],binomial),test="Chisq")
out[count,2]<-as.numeric(result[2,5])
}

for (count in 1:ncol(geno)){

##perform GC
result<-(anova(glm(formula(paste("geno[,",count,"]~",form,"pheno[,phnum]")),
binomial),test="Chisq"))
out[,3]<- gc(out[,1])
out[,4]<-gc(out[,2])

}
return(out)
}

```

R Code für Durchführung der 'Genomic Control' Korrektur

```
gc<-function(pval){
```

```

## Implements simple, fequentist genomic control (Devlin and Roder, 1999)
## on a list of P values.
## Chisq distribution recovered via a normal distribution.
## Control based on all data. (Maximum control of false positive;
## maximum loss of power).
##
## Written by J. White, NIAB, Huntingdon Rd, Cambridge, CB3 0LE.
## Jon.white@niab.com.
##
## Version 1.1.
## 28-01-09
## No guarantee offered or implied.
##
  lambda<-median(qnorm(pval/2, mean = 0, sd = 1, lower.tail = F,
log.p = FALSE)^2)
lambda<-lambda/.455
if(lambda>1){
chi.gc<-(qnorm(pval/2, mean = 0, sd = 1, lower.tail = F, log.p =
FALSE)^2)/lambda
return(pchisq(chi.gc, df=1, ncp=0, lower.tail = F, log.p = FALSE))}else{
return(pval)}
}

```

5.7 Visuelle Darstellung der Ergebnisse aus R 2.8.1

Die folgenden Abbildungen zeigen die Graphik 'S.Association.GC.P' (C) der Berechnungen der Marker-Eigenschafts-Assoziationen mit R. 2.8.1. Die Zusammenfassung dieser Berechnungen ist in Tabelle 3.9 zu finden.

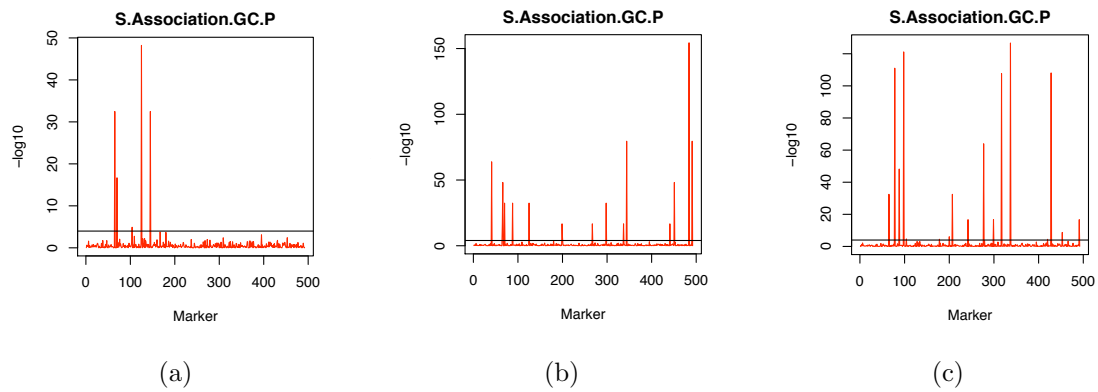


Abbildung 5.3: Ergebnisse aus R 2.8.1 für Anthocyangehalt im Blatt; (a) K=10, (b) K=11, (c) K=12

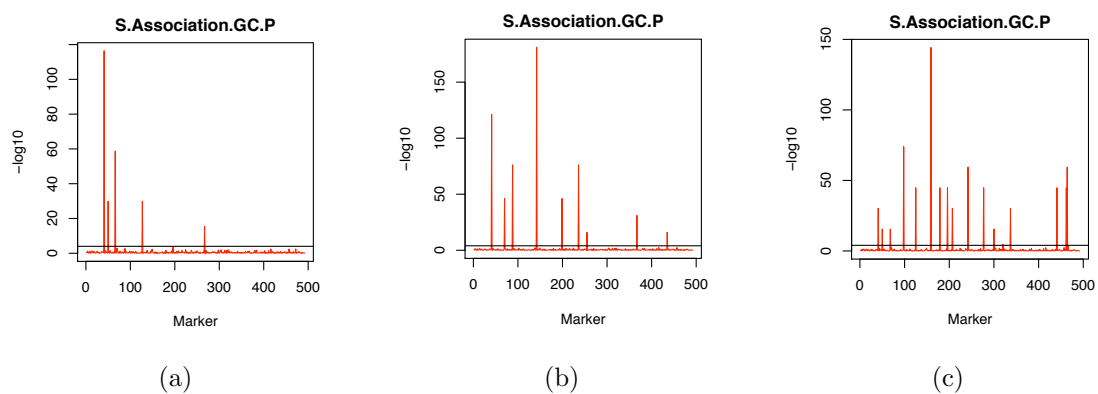


Abbildung 5.4: Ergebnisse aus R 2.8.1 für Blütenfarbe; (a) K=10, (b) K=11, (c) K=12

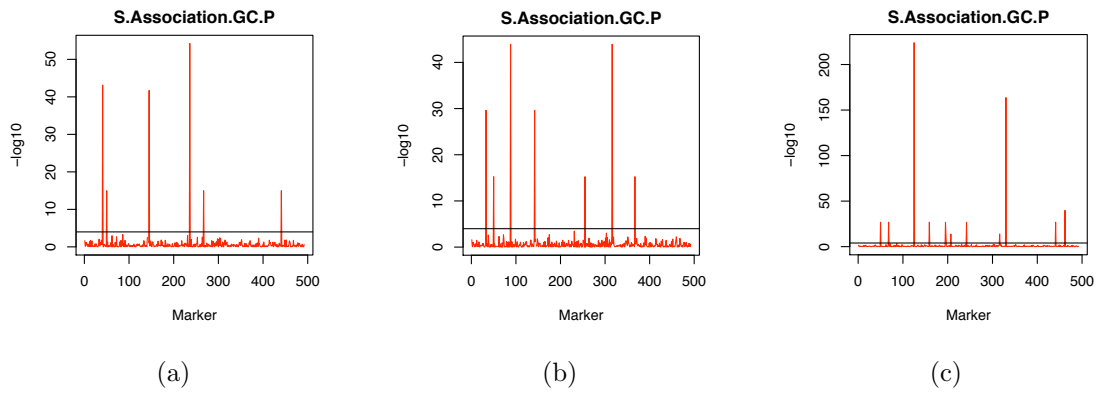


Abbildung 5.5: Ergebnisse aus R 2.8.1 für Blütenkonsistenz; (a) K=10, (b) K=11, (c) K=12

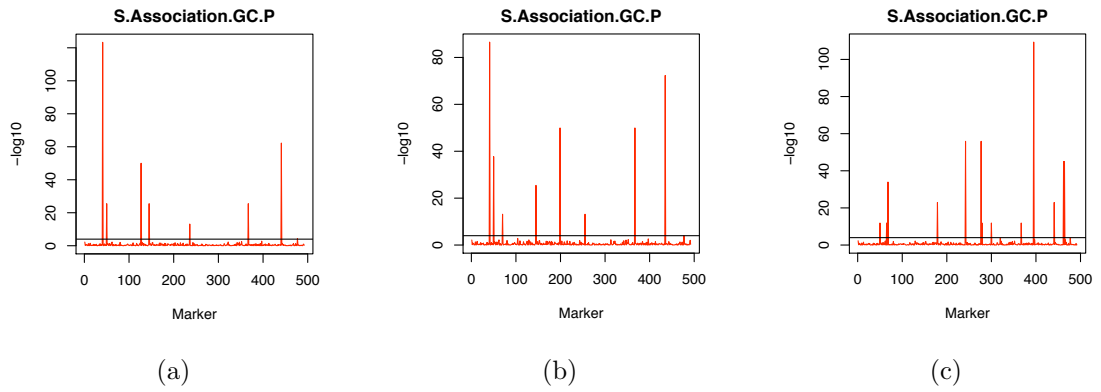


Abbildung 5.6: Ergebnisse aus R 2.8.1 für Blütengröße; (a) K=10, (b) K=11, (c) K=12

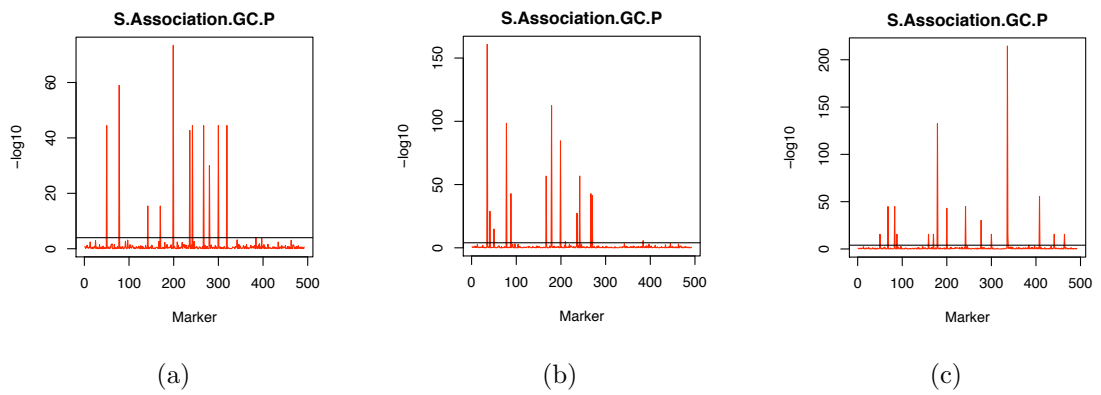


Abbildung 5.7: Ergebnisse aus R 2.8.1 für Blütentyp; (a) K=10, (b) K=11, (c) K=12

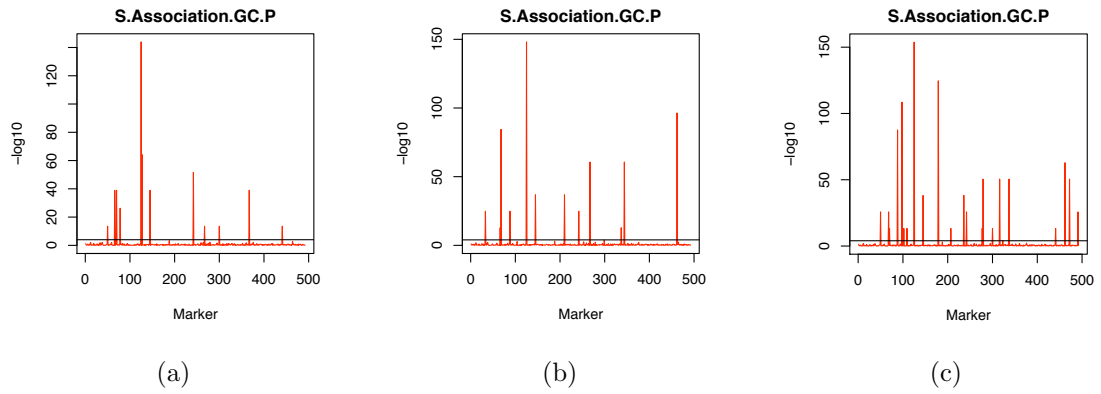


Abbildung 5.8: Ergebnisse aus R 2.8.1 für Duft; (a) K=10, (b) K=11, (c) K=12

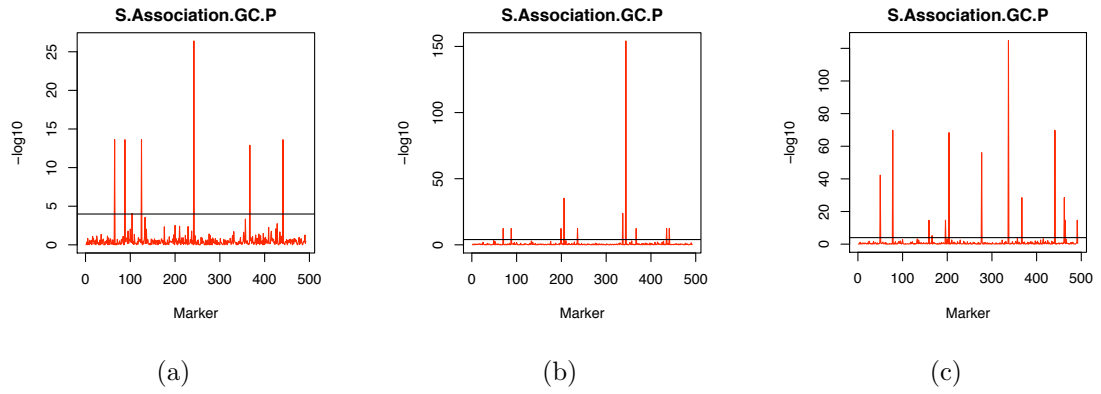


Abbildung 5.9: Ergebnisse aus R 2.8.1 für Blattmuster; (a) K=10, (b) K=11, (c) K=12

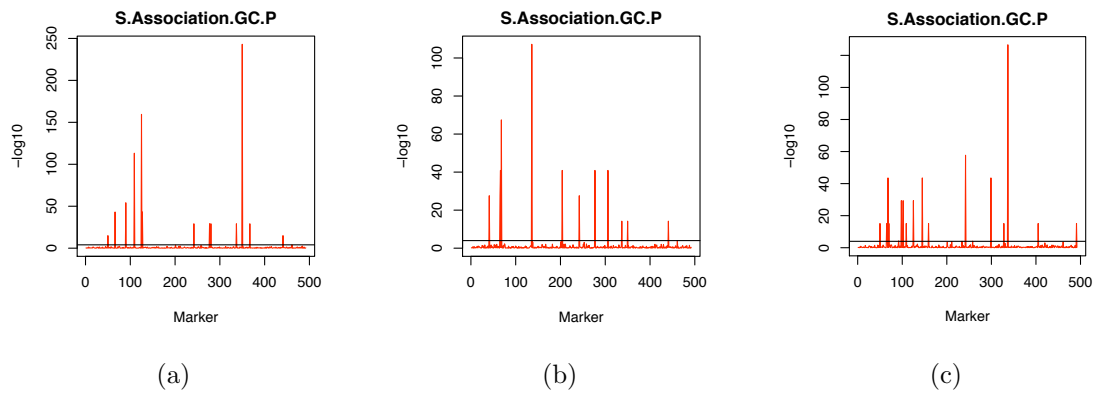


Abbildung 5.10: Ergebnisse aus R 2.8.1 für Blattform; (a) K=10, (b) K=11, (c) K=12

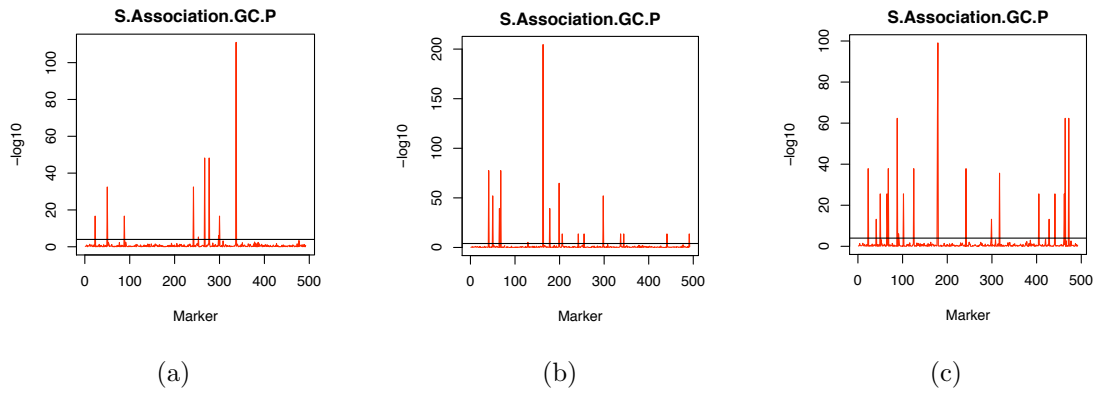


Abbildung 5.11: Ergebnisse aus R 2.8.1 für Blatttyp; (a) K=10, (b) K=11, (c) K=12

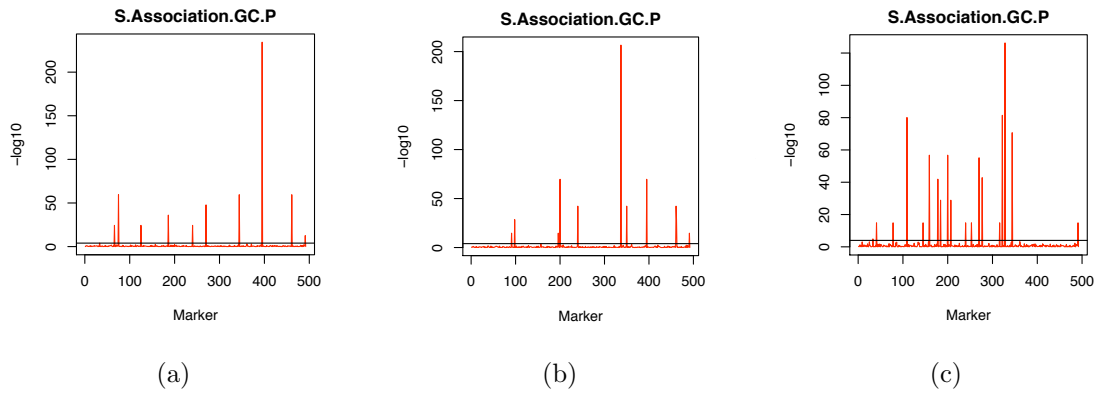


Abbildung 5.12: Ergebnisse aus R 2.8.1 für Herkunft; (a) K=10, (b) K=11, (c) K=12

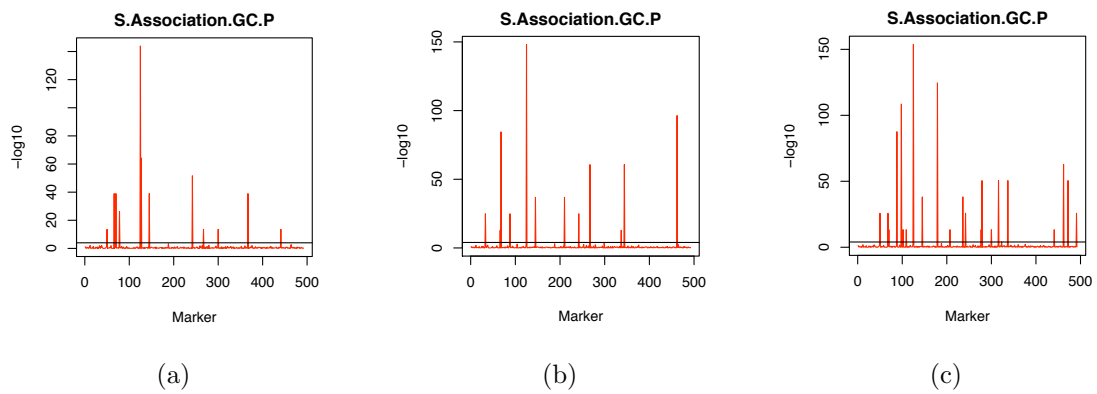












Abbildung 5.13: Ergebnisse aus R 2.8.1 für Rispenlänge; (a) K=10, (b) K=11, (c) K=12








5.8 Genotypensammlung

Tabelle 5.15: Gruppe 4

Klonnummer	Name	Haupt- Untergruppe	Nummer in Structure	Herkunft	Bild
19	unbekannt	4	1	unbekannt	
5700	Koe	4	130	N	
0738	Madonna	4	5	unbekannt	
0731	Grace	4	10	unbekannt	
1504	Wakanda	4	20	B	
0700	Nayeli	4	38	C	
1548	unbekannt	4	104	M	
1549	Talutha	4	105	M	
0754	Evita/Tadewi	4	44	C	
1533	Orenda	4	101	unbekannt	
1506	Kimama	4	91	B	
1537	unbekannt	4	103	unbekannt	
0003	<i>Ph philipiensis</i>	4	136	unbekannt	
5702	unbekannt	4	30	N	
0006	<i>Ph stuartiana</i>	4	139	unbekannt	









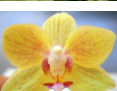
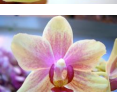



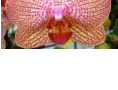
Bildquelle (wenn nicht anders angegeben): <http://www.bockbioscience.com/>

Tabelle 5.16: Gruppe 1A

Klonnummer	Name	Haupt- Untergruppe	Nummer in Structure	Herkunft	Bild
21	unbekannt	1A	2	unbekannt	
1358	Marita	1A	144	H	
0890	unbekannt	1A	75	unbekannt	
1529	unbekannt	1A	100	unbekannt	
0892	unbekannt	1A	77	A	
1511	A Chorus Line	1A	93	C	
1502	unbekannt	1A	90	B	
1534	West Side Story	1A	102	unbekannt	
0744	Pauline	1A	3	C	
0739	Roxanne	1A	11	C	
0860	Musette	1A	69	C	
5710a	unbekannt	1A	31	unbekannt	
5710b	unbekannt	1A	131	P	













Bildquelle (wenn nicht anders angegeben): <http://www.bockbioscience.com/>

Tabelle 5.17: Gruppe 2A

Klonnummer	Name	Haupt- Untergruppe	Nummer in Structure	Herkunft	Bild
0868	Roulette	2A	70	C	
1569	Pearly	2A	34	O	
1582	Memory	2A	116	C	
1499	Marit	2A	19	unbekannt	
0841	Sandrina	2A	59	C	
0839	Gilda	2A	58	C	
0826	Pamina	2A	57	C	
0836	Salome	2A	7	C	
1583	Diadem	2A	117	D	
1584	Sirius	2A	118	D	
1477	Sarah	2A	9	D	
3259	unbekannt	2A	25	L	
0727	Mimi	2A	40	C	
0843	Aida	2A	60	C	

Bildquelle (wenn nicht anders angegeben): <http://www.bockbioscience.com/>

Tabelle 5.18: Gruppe 3

Klonnummer	Name	Haupt- Untergruppe	Nummer in Structure	Herkunft	Bild
1535	unbekannt	3	21	unbekannt	
0007	<i>Ph gigantea</i>	3	140	unbekannt	
1307	Liodoro	3	4	F	
0005	<i>Ph cornu cervi</i>	3	138	unbekannt	
0008	<i>Ph florensis</i>	3	141	unbekannt	
0004	<i>Ph amboinensis</i>	3	137	unbekannt	
0001	<i>Ph violaceae</i>	3	134	unbekannt	
0011	<i>Ph violaceae</i>	3	145	unbekannt	
0002	<i>Ph hieroglyphica</i>	3	135	unbekannt	
0012	<i>Ph hieroglyphica</i>	3	146	unbekannt	
0009	<i>Ph mariae</i>	3	142	unbekannt	
0010	<i>Ph pulchra</i>	3	143	unbekannt	














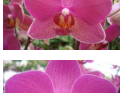

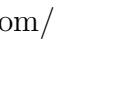
¹ <http://www.orchidspecies.com>² <http://www.shop.schwerter-orchideenzucht.de>³ <http://www.orchideen-lucke.de>Bildquelle (wenn nicht anders angegeben): <http://www.bockbioscience.com/>

Tabelle 5.19: Gruppe 2C

Klonnummer	Name	Haupt- Untergruppe	Nummer in Structure	Herkunft	Bild
1331	Limonella	2C	79	unbekannt	
1574	Backgammon	2C	112	O	
1570	Barricade	2C	110	O	
1573	Ethan Freeman /Mikado	2C	111	O	
1572	unbekannt	2C	35	O	
0866	Scrabble	2C	33	C	
0823	Nabucco	2C	15	C	
0897	Soraya	2C	78	D	
0821	Goldie	2C	56	C	
0814	Whitney	2C	6	C	
0829	Marie	2C	32	C	
4001	Raissa	2C	132	unbekannt	
3216	unbekannt	2C	133	L	
1536	The Lion King	2C	22	B	
1581	Banana Split	2C	36	R	
0809	Tosca	2C	50	C	
0844	Xerxes	2C	61	C	










Bildquelle (wenn nicht anders angegeben): <http://www.bockbioscience.com/>

Tabelle 5.20: Gruppe 2B (Teil 1)

Klonnummer	Name	Haupt- Untergruppe	Nummer in Structure	Herkunft	Bild
0849	Carmen	2B	8	C	
1498	Sarema	2B	18	unbekannt	
5129	unbekannt	2B	37	unbekannt	
0852	Jacqueline	2B	65	C	
0858	Sarasto	2B	68	C	
0813	Boheme	2B	52	C	
0816	Troubadour	2B	53	C	
0850	unbekannt	2B	64	C	
0855	Mephisto	2B	16	C	
0845	Laisa	2B	62	C	
0853	Jeanne D'Arc	2B	66	C	
0857	Alexandra	2B	67	C	
0774	Alexa	2B	13	C	
0804	Felipa	2B	49	C	
1420	Magic Dream	2B	80	C	
1421	Magic Love	2B	81	C	





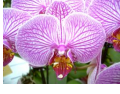




Bildquelle (wenn nicht anders angegeben): <http://www.bockbioscience.com/>

Tabelle 5.21: Gruppe 2B (Teil 2)

Klonnummer	Name	Haupt- Untergruppe	Nummer in Structure	Herkunft	Bild
1422	Magic Touch	2B	82	C	
1423	Magic Moment	2B	83	C	
1424	Magic Kiss	2B	84	C	
0870	Letizia	2B	71	J	
1577	unbekannt	2B	113	L	
1564	unbekannt	2B	108	L	
1552	Cherry Drops	2B	106	C	
1580	unbekannt	2B	115	L	
1560	Beyonce	2B	107	C	
1599	unbekannt	2B	121	L	
3247	unbekannt	2B	127	L	
0848	Don Giovanni	2B	63	C	
0819	Lulu	2B	54	C	
0820	Don Carlos	2B	55	C	











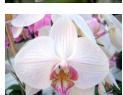


Bildquelle (wenn nicht anders angegeben): <http://www.bockbioscience.com/>

Tabelle 5.22: Gruppe 1C (Teil 1)

Klonnummer	Name	Haupt- Untergruppe	Nummer in Structure	Herkunft	Bild
3200	unbekannt	1C	23	L	
0801	unbekannt	1C	48	M	
3219	unbekannt	1C	24	L	
1475	Eugenia	1C	86	D	
0728	Idomeneo	1C	41	unbekannt	
1471	Amelie	1C	85	D	
3213	unbekannt	1C	123	L	
0883	unbekannt	1C	73	J	
3244	unbekannt	1C	126	L	
1492	Wotan	1C	87	unbekannt	
0811	Violetta	1C	51	C	
3264	unbekannt	1C	26	L	
3258	unbekannt	1C	129	L	
0750	Ayla	1C	42	C	
1500	unbekannt	1C	88	B	
4095	Mirabella	1C	28	K	
0726	Carolina	1C	39	C	
3209	unbekannt	1C	122	L	
3220	unbekannt	1C	124	L	
0891	unbekannt	1C	76	L	
1578	unbekannt	1C	114	L	
1594	unbekannt	1C	120	L	
0880	unbekannt	1C	72	J	
3251	unbekannt	1C	128	L	
0861	Lauretta	1C	17	unbekannt	
0771	Norma	1C	12	C	

Bildquelle (wenn nicht anders angegeben): <http://www.bockbioscience.com/>

Tabelle 5.23: Gruppe 1B (Teil 2)

Klonnummer	Name	Haupt- Untergruppe	Nummer in Structure	Herkunft	Bild
0781	Ariadne	1B	14	C	
0769	Dorothea	1B	47	C	
1585	Lohengrin	1B	119	C	
4000	Sissi	1B	27	unbekannt	
1501	Albine	1B	89	B	
1524	Tabaluga und Lili	1B	98	L	
5600	unbekannt	1B	29	G	
3243	unbekannt	1B	125	L	
1568	Anouk	1B	109	C	
0888	La Dame Blanche	1B	74	unbekannt	
1520	Graciela	1B	97	L	
0768	unbekannt	1B	46	M	
0751	Cecile	1B	43	C	
0767	Donata	1B	45	M	
1510	unbekannt	1B	92	C	
1517	Floretta	1B	95	K	
1519	unbekannt	1B	96	unbekannt	
1513	Jasmina	1B	94	M	
1527	unbekannt	1B	99	K	

Bildquelle (wenn nicht anders angegeben): <http://www.bockbioscience.com/>

Lebenslauf

persönliche Informationen

Geburtsdatum: 10. Mai 1978
Geburtsort: Neuruppin, Deutschland

Ausbildung

Mai 2006 - Sept 2009	Promotionsstudium, Abteilung für Molekulare Pflanzenzüchtung, Institut für Naturwissenschaften, Leibniz Universität Hannover Promotionsthema: Markergestützte Analyse gartenbaulich wichtiger Merkmale von <i>Phalaenopsis</i> Hybriden durch QTL-Kartierung und genomweiter Assoziationsstudie (Betreuer: Prof. Dr. Th. Debener)
Okt 2000 - Mai 2006	Gartenbaustudium, Universität Hannover Abschluss: Dipl.-Ing. agr. Thema und Ort der Diplomarbeit: „Untersuchungen zur genetischen Diversität von <i>Osteospermum</i> sp. mit Hilfe molekularer Marker, Abteilung für Molekulare Pflanzenzüchtung, Institut für Naturwissenschaften, Universität Hannover (Betreuer: Prof. Dr. Th. Debener)
Aug 2003 - Jan 2004	Auslandsemester, Teilnahme am 'Hortecus' Austauschprogramm, Oklahoma State University, Stillwater, Oklahoma, USA
Aug 1997 - Juni 2000	Ausbildung zur Floristin, Bender Blumen- und Gartenkunst, Neuss, Deutschland
Juni 1997	Abitur, Gymnasium Wittstock, Deutschland

Praktika

Aug. - Sept. 2004	Abteilung für Marketing anschließend Abteilung für Forschung und Entwicklung, Selecta Klemm, Stuttgart, Deutschland
Jan. - Mai 2004	Selecta First Class, Encinitas, Kalifornien, USA
Okt. 2001 - Febr. 2002	studentische Hilfskraft, Institut für Physik, Universität Hannover

Veröffentlichungen

Gawenda I. and Debener T., Genetic diversity of *Osteospermum* genotypes analysed by AFLP and chloroplast SSR markers, European Journal of Horticultural Science, Europ. J. Hort. Sci. 74: 86-94

Borchert T. and **Gawenda I.**, Development and application of high-throughput Amplified Fragment Length Polymorphism technique in *Calluna vulgaris* (Ericaceae) Electronic Journal of Biotechnology, accepted

Konferenzbeiträge

- 14. Jan. 2009 Vortrag: 'Association studies and the development of a Linkage map in Phalaenopsis Hybrids' (W347), PAG XVII, San Diego, CA, USA
- Sept. 2007 Vortrag: 'Development of Mikrosatellite markers for *Phalaenopsis*', GPZ Konferenz, Quedlinburg, Germany
- Sept. 2006 Poster: 'Genetic diversity of *Osteospermum* cv. as revealed by AFLP markers', Plant Genetics Conference, Kiel
- Sept. 2009 Poster: Estimation of Linkage Disequilibrium and Population structure in *Phalaenopsis* hybrids, Dundee, UK

Danksagung

Ich bedanke mich recht herzlich bei Herrn Prof. Dr. Th. Debener für die Überlassung des Themas und die hervorragende Betreuung.

Bei Frau Prof. Dr. Traud Winkelmann möchte ich mich für die Übernahme des Koreferates bedanken.

Mein Dank gilt auch der Firma Bock Bio Science GmbH für die Bereitstellung des Themas und des Pflanzenmaterials. Insbesondere möchte ich mich hier bei Frau Fredericke von Rundstedt und Frau Dr. Angela Schröder-Lorenz für die gute Zusammenarbeit bedanken.

Den Mitarbeitern des Instituts für Molekulare Pflanzengenetik danke ich für die tolle Zeit und die vielen guten Ratschläge. Insbesondere bedanke ich mich bei Marcus Linde und Helgard Kaufmann für die kritische Durchsicht des Manuskripts. Bei Monika Spiller und Ann-Katrin Lühmann bedanke ich mich fürs Zuhören, Diskutieren und Lachen.

Thanks to James Cockram for the stimulating discussion and the help to reveal the 'real' population structure.

Tina, Dani und Yvonne - danke für eure Freundschaft und euren Halt.